US Army Corps of Engineers®
Engineer Research and Development Center

ERDC INNOVATIVE SOLUTIONS
for a safer, better world

Military Facilities Engineering Technology

Social and Political Event Data to Support Army Requirements

Volume 1

Timothy K. Perkins, Colin D. Wood, Raimundo F. Dos Santos Jr., William D. Meyer, Noah W. Garfinkle, Xue Wang, Susan I. Enscore, Lucas A. Selig, and George W. Calfas

November 2017

Event Datasets – Actors, Victims, Context

Mission Relevance

|	
Civil Considerations	Sociocultural Analysis
Site Selection	Political Power
Routing	Protests & Rioting
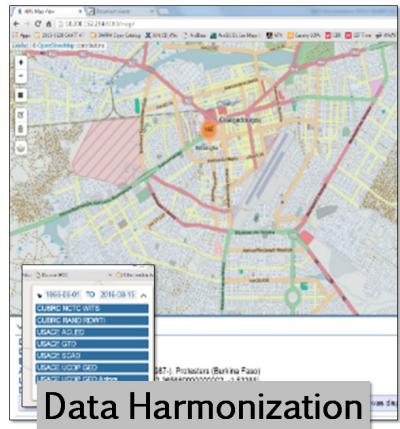
Data Harmonization
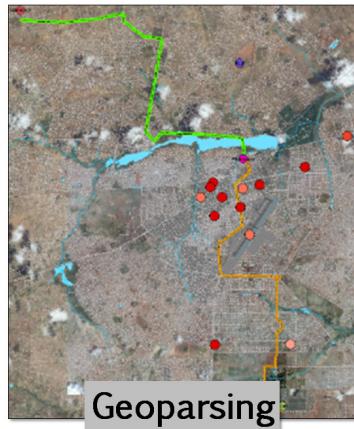
Geoparsing
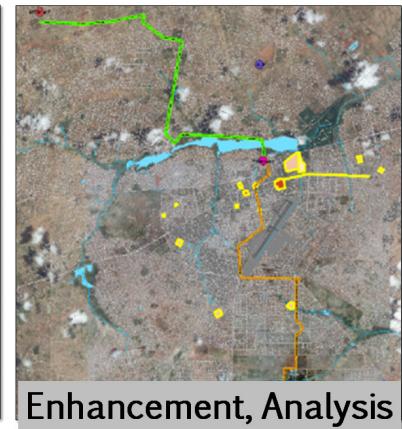
Enhancement, Analysis

Source of protest photos:
https://www.washingtonpost.com/bolgs/monkey-cage/wp/2014/11/02/burkina-fasos-uprising-part-of-an-ongoing-wave-of-african-protests/

The U.S. Army Engineer Research and Development Center (ERDC) solves the nation's toughest engineering and environmental challenges. ERDC develops innovative solutions in civil and military engineering, geospatial sciences, water resources, and environmental sciences for the Army, the Department of Defense, civilian agencies, and our nation's public good. Find out more at www.erdc.usace.army.mil.

To search for other technical reports published by ERDC, visit the ERDC online library at http://acwc.sdp.sirsi.net/client/default.

Social and Political Event Data to Support Army Requirements

Volume 1

Timothy K. Perkins, Colin D. Wood, William D. Meyer, Noah W. Garfinkle, Susan I. Enscore, Xue Wang, Lucas A. Selig, and George W. Calfas

Construction Engineering Research Laboratory
U.S. Army Engineer Research and Development Center
2902 Newmark Drive
Champaign, IL 61822

Raimundo F. Dos Santos Jr.

U.S. Army Engineer Research and Development Center
Geospatial Research Laboratory
ATTN: CEERD-PA-A
Cude Bldg. 2592
7701 Telegraph Road
Alexandria, VA 22315-3864

Final report

Approved for public release; distribution is unlimited.

Abstract

Military success requires applying judgement and decision making in a high-tempo atmosphere, based on available information. Geographic data at the city level is not enough spatial fidelity for tactical-level analyses. Violent Events Socio-Cultural Analysis (VESCA) work enables an analyst to evaluate and integrate multiple data sources, work with enhanced event data spatial resolution, and analyze and/or visualize the data to produce mission-relevant information. Hand-coded datasets can be more precise, but they require added time and labor to produce, have a significant lag between last observation and present day, are produced with varying schemas, and often duplicate events across datasets. This report includes background regarding event data sources; study of protests, demonstrations, and rallies; and relevant analytical methods. It describes doctrine regarding civil considerations, sociocultural analysis, and contingency basing to present how event data can be transformed from its original form and interpreted to support doctrinal analysis. The report also describes enhancing event data through geoparsing and through harmonization processes and tools to align datasets to a common schema and identify duplicate entries. Finally, the report presents how data may be analyzed and processed for mission-relevant results. The VESCA team's work yielded an event data harmonization prototype and recommendations for refinement.

Contents

Report Documentation Page

Figures and Tables

Figures

Tables

Preface

1 Introduction

1.1 Background

The political, military, economic, social, infrastructure, information, physical environment, and time (PMESII-PT) operational variables of an AOR affect the operational mission space and thus, they affect unit success ([Army Doctrine Reference Publication] ADRP 3-0 2016). Additionally, "Army leaders filter relevant information categorized by the operational variables into the categories of the mission variables used during mission analysis. They use the mission variables to refine their understanding of the situation. The mission variables consist of mission, enemy, terrain and weather, troops and support available, time available, and civil considerations (METT-TC)" (ADRP 3-0 2016). "…Civil considerations analysis [is] focused on the factors (areas, structures, capabilities, organizations, people, and events [ASCOPE]) affecting the civil component of the AO" ([Field Manual] FM 3-57 2014).

This analysis of civil considerations is not the sole domain of civil affairs personnel, but rather the domain of all personnel who support commanders who need timely dissemination of information to develop their situational understanding in order to plan missions. This activity is a cyclical process that provides the foundation for mission operations and results in a new series of system feedback. That feedback then generates more information and new questions for the next cycle of operational planning.

This technical report describes work conducted to support Army Warfighting Challenge #1 "Develop Situational Understanding: How to develop and sustain a high degree of situational understanding while operating in complex environments against determined, adaptive enemy organizations" (ARCIC 2017).

NOTE: Portions of this report are not included in this unclassified and unlimited release. Paragraphs removed have been noted. See Volume 2, the limited distribution version of this publication, to access For Official Use Only content.

(FOUO content removed here.)

This report presents work completed under a work unit titled "Violent Events Socio-Cultural Analysis" (VESCA), which was designed to assist with overcoming the limitations in human cognition that are associated with having to assimilate vast quantities of information about potential temporal and spatial sociocultural risks present in the operational mission space at a tactical scale. Through integration of VESCA capabilities and requisite data, analysts and Soldiers at operational and tactical echelons will gain access to civil considerations information, which will contribute to more holistic picture and greater situational understanding, and thus better operational planning, analysis, and chances for mission success.

Mission Importance: Existing geospatial analysis tools are not built for the order of complexity demanded by Urban Operations (UO). Urban features contributing to analysis of maneuver, avenues of approach, fires, hazards, and communication are not well-defined and typically not collected, extracted, or stored in a manner to facilitate operational nor tactical analysis on a mission ready device. Challenges specific to UO are therefore not sufficiently addressed by geospatial analysis tools, leaving UO Warfighters more vulnerable and with less-than-optimum situational awareness in UO campaigns.

1.2 Objective

The VESCA work unit's objective was to develop and deploy a tool that contributes to efficient processing and exploitation of event data, and supports analysis of patterns from such events. The VESCA products consist of geographic information system (GIS) capabilities that will contribute to operational contingency base (CB) site selection and to tactical, multi-modal routing analyses. VESCA products thus enhance the warfighter's situational understanding of a complex urban environment.

1.3 Approach

The VESCA team accomplished its objective through a multi-step research process that included a literature review; adoption of a framework and implementation of a tool for managing violent event data; analysis of multiple sources to populate the framework; design of enhancement methods and processing methods; and integration of data and products into a notional routing scenario set in Ouagadougou, Burkina Faso, as well as into a notional CB site selection analysis tool set in Dhaka, Bangladesh.

VESCA team members analyzed peer-reviewed literature and theoretical grounds for place-based analysis and for sociopolitical events such as protests and demonstrations. VESCA team members reviewed literature describing the evolution of protests and demonstrations, as well as criminological analyses as related to their spatial and temporal context. Emphasis was placed primarily on the built environment as perceived by the population; how public and private space, inequality, and power contribute to spatial components of protests and demonstrations; and how to effectively process and encode past locations and infer potential future locations of protests and demonstrations.

Based on the literature review, VESCA team members acquired diverse event datasets and implemented a data harmonization capability to integrate data about past events (especially protests and demonstrations). VESCA continued by designing and implementing capabilities and methods to further enhance the event data with greater location detail and then, to process that data to yield geospatial, urban-level, sociocultural products that support greater situational awareness.

1.4 Scope

VESCA developed and demonstrated processes and tools to fuse violent event data from diverse sources, enabling an analyst to evaluate and integrate multiple sources of data, enhance available event data spatial resolution, and analyze and/or visualize the data. VESCA capabilities enable an analyst to generate spatial GIS products that provide data for socially and politically important events. The data records come from multiple open sources and include duplicates; VESCA capabilities enable analysts to automatically identify duplicates and align records to gain information. The records generally include date, event type, location, actor, and target, but some sources provide more information, such as event descriptions. Location details generally depend on source articles and geoparsing implementation. Articles about events sometimes include high-resolution location information; including this type of information is especially true for event data involving violent events, protests, riots, and other gatherings.

While VECSA was part of the ERDC-run Geospatial Analysis at the Tactical Edge (GATE) work package[1] through FY16, VESCA supported a GATE demonstration event and GATE transition planning by participating in a

---

[1] GATE report is in preparation by the ERDC Geospatial Research Laboratory.

routing demonstration scenario set in Ouagadougou, the capital of Burkina Faso. VESCA analyzed and used data from 1996 through 2016. VESCA contributed information on areas with past violent events and areas of symbolic importance for possible avoidance when routing tactical operations in a complex urban system. This case study focused on civil unrest expressed through protests, demonstrations, and violent events to indicate areas of sensitivity and potential instability. The research required considerable manual pre- and post-processing of event data.

VESCA event data was also used as part of another ERDC work package, Spatio Temporal Reasoning and Introspection of Data and Embedded Relationships (STRIDER). The STRIDER case study focused on acquisition of diverse event data sources, the methodology for harmonizing the event datasets to a common schema for study in a common workspace, and visual analytics within the STRIDER tool.

VESCA also supported the Engineer Site Selection for the Tactical Environment (ENSITE) work package. The ENSITE case study focused on civil unrest for the city of Dhaka, Bangladesh (Al-Chaar et al. 2016). In this study, VESCA prepared manually coded event locations for several hundred events, and the data was used in the analysis of CB site locations.

1.5 Technology transfer

VESCA produced an event data harmonization prototype, which was configured on the Army Geospatial Enterprise (AGE) Node. The event data harmonization prototype was applied to multiple event datasets, allowing divergent data schemas to be collapsed to a common data model, and duplicate events to be identified and resolved. VESCA also produced an event data geoparsing prototype to aid in enhancing the location details about events. The geoparsing prototype and data demonstrated the possibility of enhancing location extraction, but the prototype requires further development.

2　Event Data

2.1　Background

Since the early 1970s, researchers have been working to find ways in which vast amounts of information could be quickly assimilated from the predominant form of distribution (text) into quantifiable units of analysis which would convey societal stability/instability. Initial efforts in the 1970s relied primarily on large teams of humans as content analysis coders. These coders read and analyzed vast amounts of textual information regarding international topics of interest, with an eye to discerning the relationships between a country's instability and broad trends in political, social, economic, and demographic factors (O'Brien 2010).

The predominant unit of analysis targeted in these studies became known as an "event," which involves an actor, a target, a time period, an activity, and an issue around which the event revolves (Azar 1975). Another well-accepted definition of an event is given by (Gerner et al. 1994) wherein an event is an interaction, associated with a specific point in time that can be described in a natural language sentence. Here, we use the definition from Beieler et al. (2016, 98), that says political event data are

> *"…records of interactions among political actors using common codes for actors and actions, allowing for the aggregate analysis of political behaviors. These data include both material interactions between political entities and verbal statements. Such data are common in international relations, recording the spoken or direct actions between nation-states and other political entities."*

After reviewing the definition above, it can be noted that the subject and object of the definition is an element of a set of actors, and its verb is an element of a set of actions that contain transitive verbs. The quantitative analysis of event data has traditionally been meant to characterize a detailed account of interaction between countries. However, more recently it has been broadly applied to analyze behavior of intrastate actors at the regional and subregional levels (Veen 2008). In the 1990s, the U.S. National Science Foundation (NSF) launched the Data Development in International Relations (DDIR) effort (Merritt, Muncaster, and Zinnes 1993). The DDIR sought to inspire new development of innovative methods for collecting data in international relations studies. Significant improvements in

data collection could now be accomplished with the advent of digital news media over previous hard-copy forms of sociocultural information that predominantly utilized large teams of human coders. Newswire services could now be directly downloaded from the internet such as Agence France Presse, Reuters, or Associated Press and easily converted into a machine-readable format ready for processing. With digital media, sources that were readily available for processing the area of focus could be further refined to process areas of greater spatial granularity. Event data analysis prior to 1990 could only compare and contrast state on state or country on country actors but post-1990, it would be possible for an analysis to go down to a subregional or even a city level.

To derive meaning from the information collected from various newswires requires that the event being reported is structured in a specific way to represent the elements (i.e., actor, target, time period, activity) of an event description. There have been several event processing frameworks developed to accomplish this activity, some of the major ones are the Integrated Data for Event Analysis (IDEA; Bond et al. 2003), World Events Interaction Survey (WEIS; McClelland 1978; Goldstein 1992), Conflict and Peace Database (COPDAB; Azar 1993) and Conflict and Mediation Event Observations (CAMEO; Gerner et al. 2002; Schrodt et al. 2008). These processing structures also enable smoother and faster machine processing of data. The speed and magnitude of data that can now be processed by machines eclipses what was possible with human coders. Schrodt (2001) reported an average number of news articles processed per day by human coders as approximately 40, and automated coding at that time approximated 3000 events per second; "the equivalent of what a human coder does in three months" completed in one second. Several studies have also shown that there is no significant improvement in coding reliability of human coders over that of machine processing (Schrodt 2001).

The CAMEO coding schema was developed to account for changes in international conflict, moving from the traditional focus on state actors to include that of substate and nonstate actors and organizations (Schrodt et al. 2008). Recent literature identifies a number of issues addressing the current state of event data processing, problems, and promises (see Schrodt 2015; Chojnaki 2012; Weidmann 2016) stemming from a lack of "gold standard" event datasets, inconsistency in data due to over- and under-reporting, etc. One issue that remains at the forefront, however, is the "open source geocoding issue" (Schrodt 2015, 17).

In order to automate event data coding, Schrodt et al. (2008) expanded WEIS and COPDAB actor dictionaries to more accurately portray each actor involved, as well as the actions of the actors. To do so, CAMEO uses an Actor-Verb-Target relationship to code, while also gathering other pertinent information such as a generalized location. For instance, a fictional article's tagline would appear as follows:

> *"Rio de Janeiro, Brazil--Students marched on the Gustavo Capanema Palace yesterday to protest an increase in education costs by the Brazilian government."*

In this case, CAMEO would recognize "students" as the actor, "protest" as the verb, and "government" as the target—each assigned a code based on actor and verb dictionaries, and often assigned a generalized score between -10 and 10 to indicate whether the event is more conflictual (-10) or cooperative (10). While CAMEO itself does not offer a built in geocoding capability, Schrodt (2015, 17) acknowledges, "[g]eocoding probably should be integrated into the coding ontologies: not every event has a meaningful location, and assigning locations where they are irrelevant simply adds noise." Careful automated geocoding, specific to the text/event being coded, seems to be the missing link in geographically specific event data.

Geographically, most NER packages would place the fictional event as occurring in Rio de Janeiro, Brazil, as indicated in the tag line. Geocoding to the city or greater level (e.g., region or state) is most common among automated coding methods, as the geographic place name is easily located in the tag line. In this case, however, there is a more exact location of interest to us—the "Gustavo Capanema Palace" which can be located to an exact address in the city of Rio de Janeiro that most gazetteers will not readily identify and code. This, according to Schrodt (2015, 17), is still a major area in which automated event coding remains lacking, but one where the "payoffs would be huge."

The CAMEO coding schema enables machine and human coding of political event data to be replicated, updated to reflect changing actors, and used interchangeably across platforms with common codes. CAMEO's real benefit, however, is the simplicity in machine coding that is cost- and time-effective, not to mention sustainable over time (Beieler et al. 2016).

2.2   Sources

Prior to beginning the VESCA project, Army analysts identified a variety of exemplary event data sources relevant to understanding the context and significance of violent events for Army planning and operations. These data sources included the Armed Conflict Location & Event Dataset (ACLED) [2], Social Conflict in Africa Database (SCAD)[3], and the Uppsala Conflict Data Program[4] Georeferenced Event Dataset (GED). Each of these academically rigorous and traceable datasets had proven valuable to analyzing and understanding violent activity, actors involved, trends, and relationships to historical, social, economic and other factors. However, for military analyst use, these and similar datasets needed updating because the update cycle was insufficient for Army analyst use. Additionally, the datasets often lacked spatial details below the city level. In collaboration with VESCA, additional data sources were identified, including some that offered the potential for temporal update frequency sufficient for Army analysts and additional data that could yield spatial event details. Table 1 summarizes the event data sources used as part of VESCA. For more on event data sources, Yonamine (2013) provides detailed discussion and examples.

Table 1. Sources of event data.

Source Name	Acronym	Summary	Website	Note
Social Conflict in Africa Database	SCAD	"The Social Conflict in Africa Database (SCAD) includes protests, riots, strikes, inter-communal conflict, government violence against civilians, and other forms of social conflict not systematically tracked in other conflict datasets."	https://www.strausscenter.org/ccaps/research/about-social-conflict.html	Data coverage 1990–2015
Uppsala Conflict Data Program (UCDP) Georeferenced Event Data (GED)	UCDP GED	Uppsala Conflict Data Program (UCDP) records violent conflicts, with an emphasis on armed violent conflicts. There are several different datasets included in the overall UPPSALA dataset, each with its own codebook and data downloads.	http://www.pcr.uu.se/research/ucdp/program_overview/	Data coverage 1989–2014
Armed Conflict Location and Event Data	ACLED	"ACLED is the most comprehensive public collection of political violence data for developing states. These data contain information on the specific dates and locations of political violence, the types of event, the groups involved, fatalities and changes in territorial control. Information is recorded on the battles, killings, riots, and recruitment activities of rebels, governments, militias, armed groups, protesters and civilians."	http://www.acleddata.com/	Data coverage 1997–present (lag ~4 days)

---

[2] Armed Conflict Location & Event Data - www.acleddata.com

[3] Social Conflict in Africa Database - www.strausscenter.org/scad.html

[4] Uppsala Conflict Data Program - ucdp.uu.se

Source Name	Acronym	Summary	Website	Note
Global Terrorism Database	GTD	The Global Terrorism Database (GTD) was developed to be a comprehensive, methodologically robust set of longitudinal data on incidents of domestic and international terrorism.	http://www.start.umd.edu/gtd/	Data coverage 1970-2015
Integrated Crises Early Warning System	ICEWS iDATA	iDATA: "The process that allows the provisioning of the models in near real-time from a variety of international, regional, national and local new sources (over 6,000). More than 38 million multilingual news stories over the past 25 years are processed to extract [who, did-what, to-whom, when, and where] from each sentence in these stories creating a right 25-year "history of the world"."	http://www.lockheedmartin.com/us/products/W-ICEWS.html	Proprietary implementation, limited access, complete text articles available; data 1991-present
Integrated Crisis Early Warning System (ICEWS) Dataverse	ICEWS Data - Open	"Event data consists of coded interactions between sociopolitical actors (i.e., cooperative or hostile actions between individuals, groups, sectors and nation states). Events are automatically identified and extracted from news articles by the BBN ACCENT event coder. These events are essentially triples consisting of a source actor, an event type (according to the CAMEO taxonomy of events), and a target actor. Geographical-temporal metadata are also extracted and associated with the relevant events within a news article."	https://dataverse.harvard.edu/dataverse/icews	Releasable portions of ICEWS event dataset, no text articles
Phoenix Data Project	PDP	"The Phoenix dataset is a new, near real-time event dataset created using the next-generation event data coding software, PETRARCH. The data is generated using news content scraped from over 400 sources. This scraped content is run through a processing pipeline that produces coded event data as a final output. Our current settings produce roughly 3,000 coded events per day. These coded events are in the standard who-did-what-to-whom format typically associated with event data. Each event is coded along on multiple dimensions, specifically source and target actors and event type."	http://phoenixdata.org/	Open-source development implementation like ICEWS iDATA; data 2014–present
Global Database of Events, Language, and Tone	GDELT	"The GDELT Project monitors the world's broadcast, print, and web news from nearly every corner of every country in over 100 languages and identifies the people, locations, organizations, counts, themes, sources, emotions, counts, quotes, images and events driving our global society every second of every day, creating a free open platform for computing on the entire world."	http://www.gdeltproject.org/	Proprietary implementation; data available, article links provided but not full text; data 1979–present

VESCA seeks to provide analysts with the ability to exploit the spatial aspects of event data at the tactical scale, providing an up-to-the-day event history overlay. The data product which is widely accepted to have these characteristics is produced by the Integrated Crises Early Warning System (ICEWS) platform. ICEWS was developed through the Defense Advanced Research Projects Agency (DARPA) and began in 2007 for an initial 4 years and then extended for an additional 3 years to 2013. The program utilized the CAMEO event data framework to preprocess data from news-wire streams which would then be utilized in social science models to forecast and understand instability across countries under investigation. The program components include iDATA (acquires, processes, and stores the

data) as well as iTRACE (querying and analyzing news data), iCAST (instability forecasting), and iSENT (sentiment analysis and opinion propagation in social media; Ward et al. 2013). Following ICEWS, some personnel involved in its development have continued to work on PDP, an open-source effort that implements a similar processing pipeline known as PETRARCH (Python Engine for Text Resolution and Related Coding Hierarchy; Schrodt et al. 2014). Both ICEWS and PDP generate similar data products; the primary differentiator for VESCA purposes is that ICEWS has direct access to full-text news articles and the derivative event coding dating back to 1995, whereas PDP does not have the same history, nor does it provide direct article access. However, PDP does provide direct access to the scripts used to complete the event data coding. CAMEO, as implemented in ICEWS and currently in PDP, has the ability to yield information at the spatial granularity of a city, but data at this scale is not suitable for tactical mission planning.

Because the ICEWS platform applies the CAMEO framework to organize its data input, VESCA can leverage ICEWS for pre-modeling data input, thereby reducing system overlaps between ICEWS and VESCA and creating complementary capabilities.

However, because ICEWS is automated (as are other CAMEO-based systems), and it was designed primarily for forecasting, it tolerates and includes substantial noise, including large volumes of miscoded events, duplicate events, and limited actor information. As described in Armstrong et al. (2015), analysts seeking to understand complex violent events often work with multiple data sources, because within and between data sources, records might mention, refer to, or be related to one another. However, as each dataset typically has its strengths and its own schema for organizing the information, it can be difficult to gather and reason across these records. Some datasets often provide high-quality information on actors, targets, and other event information, but not cover the range of event types needed by an analyst. Generally, most event datasets examined and included rarely provide geospatial information with greater detail than city name, however the Global Terrorism Database (GTD) does (whenever detailed location information is available). However, while the ICEWS dataset doesn't automatically encode such detailed location information, news articles available with the event data do offer location details, especially for certain types of events such as protests, demonstrations, and rallies. Detailed location information was derived from a manual coding

process (described in section 4.1, Geoparsing) applied to ICEWS data. Notably, this ICEWS data began as 3,806 ICEWS event records, which actually represented 892 events at 2,144 locations. In the original ICEWS datasets, all events were coded to the city level of detail, while manual coding provided evidence that more detailed location information for these events is available in article text.

Some analysts prefer or must work with peer-reviewed human-coded datasets (e.g., ACLED, SCAD, UCDP, GTD), but they need to update those datasets. Through data integration methods, analysts may take advantage of the qualities of human-coded datasets and merge these with up-to-date and full-text access datasets (e.g., ICEWS). VESCA capabilities aim to provide analysts with the ability to take advantage of the strengths of diverse event datasets to achieve greater sociocultural and place-based understanding.

2.3 Spatial components of protests, demonstrations and rallies

Appendix B provides a detailed literature review summarizing research into how space and place relate to protests, demonstrations, and rallies. As part of VESCA work, the literature review served to identify analytical frameworks adopted in social science; the meaning and role of place and space as related to protests, demonstrations, and rallies; and to identify those elements that may serve as attractors and detractors—spatially and temporally—for such events. Provided below are listings of built environment elements gathered from the literature. These lists are not definitive, but they serve to enable efficient investigation of the specific roles of these elements and other elements that may also be involved in these types of political, economic and social events. Of particular note is the fact that many of the attractor elements can be altered by the authorities to become detractor elements that are intended to prevent or discourage protests, rallies, and demonstrations.

2.3.1   Attractors in the built environment

2.3.1.1 Spatial attractors

- Large central commercial sites
- Dense multistory apartments
- High levels of marginalized populations concentrated in particular areas

- Spatial patterns and routines that are not conducive for community policing
- Large number of people in a particular place
- Large open spaces at intersections of main transitways
- Public squares or plazas
- High-level government buildings (palaces, parliaments, police/military headquarters, political party headquarters, embassies, etc.)
- High-level private buildings (corporate headquarters, banks, stock exchanges, elite residential areas, etc.)
- Historical or religious sites or centers
- Familiarity with the protest space
- Familiarity with transit routes and ease of access
- Linkage between features/protest routes
- Sidewalks or walkways that are open and accessible to pedestrians
- Open public land such as parks, playgrounds, and parking lots
- Places that provide physical access to directly confront the symbols of authority

2.3.1.2 Temporal attractors

- Low time-distance costs
- Times fitting mass transit schedules
- Times when the group is already present near the protest space
- Protests occurring at regular intervals or schedules

2.3.2 Detractors in the built environment

2.3.2.1 Spatial detractors

- Low-density residential or individual unit
- Improvised barricades or borders
- Small- and medium-sized streets defendable against protests
- Subdivided public areas (e.g., fenced off, barricaded, policed)
- Wide central boulevards as "no man's land" (hard to cross, easy to police)
- Large public squares and other spaces that can be "filled" with street furniture (benches, bollards, fountains, planters, etc.) to inhibit large crowds
- Space too constrained, either by physical borders or by barriers erected on the site
- Space without strong symbolic elements of authority

- No linkages between protest spaces
- Streets with police roadblocks to turn back protesters
- Former public space that has been privatized and controlled (e.g., residential areas, parks, walkways)

2.3.2.2 Temporal detractors

- Inconvenient times for travel to protest site
- Times when possible participants are not in the area
- Infrequent mass transit schedules
- High time-distance costs

3 Mission Relevance of Event Data

As described in section 2.2 of this report ("Sources"), a broad range of data sources provide political event data, which include events such as terrorist attacks, coups, violent protests, demonstrations and other types of events. Such event data generally make reference to participating entities, include features such as an event type (e.g., protest, demonstration, terrorist attack), characterize entities as either actors and/or targets, the date on which an event occurred (or when an event began and finished occurring, depending on the data source), and some location information (e.g., the country, or perhaps city, or even sometimes specific detailed coordinates of a place within a city). Event data, when combined with other information, offers the possibility for analyses that could contribute to better political understanding of the relationships of social groups, place, and narratives that are invoked or resonate with certain population segments, social movements, power dynamics, and other topics. Even absent integration with other information, event data offers potential value.

NOTE: Portions of this chapter are not included in this unclassified publication; paragraphs removed have been noted. See Volume 2, the limited distribution version of this publication, for FOUO content.

3.1 The importance of situational understanding for contingency base site selection

(FOUO content removed here.)

(FOUO content removed here.)

Base Camps (ATP 3-37.10) further identifies the G-9/S-9 to advise the base camp commander/Battlefield Operating System-Installation (BOS-I) on "the military operations effect on civilians in the AO relative to the complex relationship of civilians with the terrain and institutions over time." (1-19). While this stresses CA study and analysis during ongoing operations (mission-specific), we can see that such activity occurs throughout the MDMP, and indeed should apply to site selection considerations with regard to contingency basing. Likewise, it notes that the "intelligence section serves as the principal staff for providing intelligence to support current and future operations and plans. …[gathering and analyzing] information on enemy, terrain, weather, and civil considerations for the base

camp commander/BOS-I." (1-18) This places both CA and intelligence activities in the fore for creating situational awareness for both commanders and planners.

3.1.1 Site selection

Base camp site selection occurs "during mission analysis/problem framing with the identification of suitable and unsuitable areas... primarily determined on an analysis of terrain and civil considerations." (ATP 3-37.10, B4–B-5) In order to optimize the selection of a site for a base camp, one must "balance between operational, sustainment, and construction requirements," and consider the "operational and mission variables" (ATP 3-37.10, 2-10) of METT-TC and ASCOPE/PMESII. While this seems intuitive and easy enough to say, how to incorporate civil considerations meaningfully into the site selection process may not be quite so.

(FOUO content removed here.)

3.1.2 Civil considerations

Civil considerations are, simply, the nonmilitary factors (areas, structures, capabilities, organizations, people, and events [ASCOPE]) affecting the civil component within the operational environment that aid the commander in understanding the effect of such variables on the mission (FM 3-57, 1-4 – 1-5; ATP 2-01.3, 3-6). Civil Affairs and intelligence personnel have developed their own methodology for assessment, though there exists no true standard for minimum requirements of information within the ASCOPE framework. Rather, such products are highly tailored—created to support the commanders' needs and fill information gaps (ADRP 2-0, 5-3).

(FOUO content removed here.)

Here, we are primarily concerned with events as our unit of analysis with regard to civil considerations and CB/base camp citing. Doctrine identifies events as "routine, cyclical, planned, or spontaneous activities that significantly affect organizations, people, and military operations" (ATP 2-01.3, 4-34). Civilian events (such as elections, riots, evacuations, etc.) can have a tremendous effect on military operations, just as military operations (such as a combat mission during a contingency operation) can have both posi-

tive and negative effects on a civilian population (FM 3-57, 4-10). Thorough analysis of political event data occurring within the AOR provides necessary insight for planners in understanding the historical and ongoing issues affecting the populace, as well as providing a means of prediction for how the people may react to a CB site.

An issue of concern that must be addressed is establishing why planners and commanders should pay attention to historical political violence with regard to the placement of a contingency base. An examination of open-source data—such as from the Integrated Crises Early Warning System (ICEWS), for example—can reveal a great deal not only about types and quantities of political violence enacted over a period of time and space, but also about those involved and, ostensibly, some level of understanding as to why the actions take place. As discussed later, the Conflict and Mediation Event Observations (CAMEO) coding schema used in ICEWS codes events in the manner of actor did action to target, generally with some additional spatio-temporal information. This allows the identification of specific groups of interest as well as possible motivating factors precipitating violence. As a product informing IPB and thus the MDMP/BCDP, inclusion of such open source material is necessary to attain greater situational understanding both of the operational environment and how CBs/base camps can help or hinder U.S. forces engaged in contingency operations.

For instance, if offensive contingency operations are to begin in an area that has historically seen a great deal of violence against the government by rebel forces, knowing where those forces have previously focused their attacks and what forces were involved could help planners understand where "better" potential CB sites might be. Likewise, understanding if those rebel forces have previously attacked U.S. civilian or government interests as well as knowing their propensity to collaborate with and/or their acceptability to the surrounding population would also inform planners. Much of this information can be gleaned from analyzing events.

3.1.3   Use case

While VESCA data can provide Geospatial Intelligence Analysts a great deal of information with regard to civil considerations in contingency basing, it is most useful as a product informing the IPB process. The utility of VESCA event data is in the enhancement of geo-located event data. Such enhancement of the existing data allows intelligence analysts the ability to

better understand the operational environment and provide greater situational understanding to the planners and commanders through the IPB process. Given that IPB is defined doctrinally as, "the systematic, continuous process of analyzing the threat and environment in a specific geographic area," (ATP 2-01.3, 1-1) event data can offer a great deal of information to planners and commanders on the history of political violence in the area of operations which may directly impact the warfighter. A notional use case is outlined below.

Users/Actors: Intelligence Analyst (35F); Command Staff (AFRICOM).

Scenario: An Intelligence Analyst (35F) assigned to AFRICOM (G-2) is tasked with examining a large, urban area in an area of regard to assist planners by providing information that is of great import to the opening of combat operations to remove an anti-U.S. regime. Specifically, planners are concerned with a particular section of the capital city where they anticipate heavy fighting and difficulty in maintaining appropriate logistics, and thus wish to establish contingency basing within the proposed AO. The analyst has access to readily available open-source event databases as well as open-source and classified information on groups, leaders, and the HN government.

The analyst begins the process by determining what data are available that may assist him in assessing the OE and providing actionable intelligence to planners. In order to do so, he begins by establishing the baseline for political violence in the area, using a variety of open-source (ACLED[5], GTD[6], UCDP[7], etc.) and proprietary/classified (ICEWS[8], SIGACT[9] if applicable, etc.) political violence databases. Through exploratory data analysis at the city-level, the analyst is able to identify actors who often oppose the government, and thus may be amenable to U.S. forces and operations in the AO, as well as counter-opposition groups which will likely resist U.S. forces and operations in the AO. The analyst compiles thorough profiles on the many groups, and provides the city-level data as their addition to the IPB product provided to commanders ahead of operations. Command and

---

[5] Armed Conflict Location & Event Data - www.acleddata.com

[6] Global Terrorism Database - www.start.umd.edu/gtd/

[7] Uppsala Conflict Data Program - ucdp.uu.se

[8] World-Wide Integrated Crisis Early Warning System - www.lockheedmartin.com/us/products/W-ICEWS.html

[9] Significant Activities reports

staff are able to examine the likelihood for violence among groups at the city level, but are unable to pinpoint more specific areas without better data.

While group profiles are helpful, commanders and planners cannot discern where, within the city, the groups are primarily operating based on the targets hit. If, however, the same data were geospatially enhanced as VESCA pursues, the same analyst would be able to begin mapping out areas at the neighborhood to facility level, enabling situational understanding of the AO that is unmatched with currently automated OSINT collection. Now, the analyst is able to provide the location of opposition headquarters that may be amenable to U.S. forces, as well as areas more likely under their control, determine areas that are more often attacked by pro-government forces, and more. Likewise, areas firmly controlled or influenced by anti-opposition and government forces can be readily seen.

See Appendix A of this report for excerpts, quotes, and other material referenced in this section.

3.2 Sociocultural analysis and the Army

In addition to direct support to contingency base siting, political event data is an important contributor to other analytical requirements of the Army. The Global Cultural Knowledge Network (GCKN) of U.S. Army Training and Doctrine Command (TRADOC) G-20 published "Socio-Cultural Analysis Framework (SCAF): A U.S. Army Guide on How to Research and Write Socio-Cultural Analyses" (GCKN 2016). The SCAF is derived from a diverse range of frameworks that had been published in approximately 84 military publications such as field manuals, training publications, handbooks, and others. The SCAF offers a taxonomic approach to arranging several related terms, descriptors, and indicators and then, associates them all back to a modified PMESII-PT framework. The SCAF framework indicates where certain sociocultural information can contribute to Army missions. To aid the team in applying the SCAF framework, the team adopts the following definition of political system and power from the Political Military Analysis Handbook (U.S. Army 2008, 3-2): "A political system is any grouping of primarily civil roles and institutions, both formal and informal that exercises authority or rule within a specific geographic boundary or organization through the application of various forms of political power and influence."

Event data could contribute to a wide range of SCAF elements especially in combination with other data, but VESCA focuses on those SCAF elements most supported independently by event data with little or limited addition of other data. Table 2 and Table 3 present a selection of SCAF excerpts including the SCAF domain and question. The right-hand column provides a brief description of how event data may be processed and used in order to address or support each of the entries.

Table 2. Excerpts of sociocultural analysis framework questions
related to event data (ERDC-CERL).

Domain	Question	Potential Processing and Contribution
Political	*Who exercises political power and how can this be measured?*	Event datasets such as ACLED, SCAD, ICEWS and others described in section 2.1 (Sources of Event Data) are prepared with key actors and/or groups identified. Combining these event datasets and extracting the actor/group information provides a means of organizing information on some of those actors/groups who exercise political power. In addition to identifying the actors/groups, many of the event datasets provide event types and/or event categories, along with a temporal dimension, which may be processed to graph and/or map indicators of political power.
Political	*Which institutions wield power? Particular social structures (tribes, clans, etc.)? Religious entities? Labor unions? Political parties? Courts? Criminal organizations?*	See above.
Political	*Are certain non-governmental organizations more powerful than others in the community or society? For example, do religious groups hold more persuasive influence over the population?*	The actors/groups tracked in the ICEWS dataset, and possible to implement with PETRARCH, include various and configurable categories such as police forces, judiciary, military, insurgents, political opposition, rebels, agriculture, business, criminal, development, education, environmental, religious, etc. Using the actor/group categories enables analyses of the organizations and others involved in political events reported in tracked media. Such study may reveal which organizations are wielding power, through which types of events and with what targets. Further analysis of related events, perhaps such as an attack followed by protests, or qualitative analysis of articles, may provide insight about population reaction.

Domain	Question	Potential Processing and Contribution
Political	*What friction points exist within the political system that has the potential to polarize society? Are there religious and civic groups who actively oppose each other's policies? Each other's sociopolitical objectives?*	Using temporal event datasets with data on actor, action and target (e.g., ICEWS), analysts may examine whether events between groups recur periodically, or whether events may be unusual or an anomaly. Thus, event data may help an analyst determine whether groups are likely to engage in conflict or cooperation. Further analysis of articles about the events may reveal specific group objectives. The ICEWS event dataset includes a 'Goldstein score' (Goldstein 1992, Schrodt 2014) associated with events, which enables an analyst to efficiently filter event types along a spectrum of conflict and cooperation.
Political	*Does the country generally have a positive or negative relationship with other countries, such as the U.S., Russia, China or Iran or other UN members?*	Event datasets include groups from within a country, and as related to other countries and actors in other countries, especially as related to state-level actors (e.g., President, official groups, etc.). Querying event datasets for an event 'source' (who took the action) of an actor from country X may yield events over time as related to many other countries, and yield events that are cooperative, or conflicting, or trending from one to the other over time.
Political	*Is there a political tradition regarding the peaceful or violent transfer of power?*	Event datasets examined under VESCA do not yet clearly address electoral events. Event datasets exist that already track coups d'états and other state-level electoral violence (e.g., Mass Mobilization Database) that could be incorporated. Additionally, emerging event coding capabilities under the PLOVER[10] program aims to extend CAMEO like coding to "contexts such as disease, natural disaster, elections, parliamentary processes and cyber-security." Until PLOVER is ready for adoption, an analyst may examine event trends surrounding dates for elections and transfer of power, but these dates must generally be acquired separately through existing databases.
Political	*How does the population demonstrate dissent?*	Several of the event datasets (e.g., SCAD, and those derived from CAMEO coding) include a range of event types that represent forms of dissent. An analyst may filter event data to those event types to determine patterns and trends related to groups that participate in particular types of events. Applying the risk terrain modeling approach described in Chapter 3 and utilizing geographically enhanced data on riots, demonstrations and protests, an analyst may understand the specific locations of such events and possible local attractors and detractors relevant to such events.

---

[10] https://github.com/openeventdata/PLOVER

Domain	Question	Potential Processing and Contribution
Political	*Does religious/ethnic/tribal identity affect political participation?*	By examining event datasets organized by actors/groups, an analyst may discern patterns of event involvement as aligned with particular identity groups. By applying the risk terrain modeling approach described in Chapter 3, and analysis of the spatial patterns of protests, demonstrations, and rallies in their area of operations, an analyst may identify contributing attractors and detractors, such as affiliation groups.
Security	*Who are the relevant coercive groups in the AO? (The SCAF defines coercive groups as those, "that have the potential to affect security policy" and may include internal or external groups, using force, threatened force, or no force.)*	By filtering event data records to actors opposing or supporting government actors, security forces, and other appropriate groups, an analyst may derive a list of potentially coercive groups.
Security	*How do coercive groups diverge or converge with local, national, regional, international, and/or U.S. agendas?*	An event list filtered to coercive groups can provide an analyst with information about how and when those groups have taken action. By examining articles about those events, or by assessing actions taken by the U.S. and its allied actors, an analyst may assess whether the coercive groups diverge or converge with others' agendas.
Security	*What is the relationship between the coercive group and the [host nation] HN Government?*	An analyst may query the event data to examine event records, should they exist, that include both the coercive group and HN Government. Additionally, an analyst may query event data to determine if the coercive group and HN Government conduct similar or related actions towards common or affiliated targets, or are both targets from common or affiliated sources.
Security	*What are the cooperative links between coercive groups (who has access/rapport/trust with whom)?*	See above.
Security	*What are the frictions between coercive groups? What is the basis of these frictions? What are the effects of these frictions?*	See above.
Cultural	*What conflicts exist between religions?*	See above, but with filtering event data records to religious actors/groups.

Domain	Question	Potential Processing and Contribution
Geographic	*What are the significant historical and religious sites in the AO?*	Event datasets described in section 2.1 will not currently reveal sites with historical or religious significance. However, by enhancing event data geographic locations through geoparsing, an analyst may determine if events recur in specific places repeatedly. By using additional event data, such as actor type, or article text, an analyst may be able to determine the significance of the event locations.

Event data sources offer analysts the opportunity to quickly derive relevant sociocultural information with minimal search and processing requirements. However, it is important to also note that event datasets are incomplete and many include noise. Leetaru (2010), Shellman (2008), and Schrodt (2001) indicate that source bias and coverage is a factor when event datasets rely on news media sources, which include datasets described in section 2.2. Such bias may be mitigated, according to Leetaru (2015) and Shellman (2008), by using diverse sources when processing media reports about events rather than relying on singular or few sources. Similarly, analysts using event data may mitigate incomplete or biased data by using multiple event datasets.

Another challenge of using event data sources to support the analyses described in the current and preceding section is that the actor and group dictionaries that underpin them are always evolving and incomplete. Thus, in such cases, an actor or group may at one point be coded generically as a rebel group, and later as a specific rebel group. With direct access to article text, analysts have the opportunity to discern more information about events and the actors and groups involved.

3.3   Risk Terrain Modeling and international relations

Risk Terrain Modeling (RTM) is described by Caplan and Kennedy (2011, 11) as "an approach to spatial risk assessment that utilizes a geographic information system (GIS) to attribute qualities of the real world to places on a digitized map. …Risk terrain maps show places where conditions are conducive for certain events to occur in the future based on the environmental context for criminogenmesis." Kennedy and Van Brunschot (2009, 4) define risk assessment as "a consideration of the probabilities of particular outcomes," whereas the UN defines it as a "methodology to determine the nature and extent of risk by analyzing potential hazards and evaluating existing conditions of vulnerability that together could potentially harm

exposed people, property, services, livelihoods and the environment on which they depend" (UN/ISDR 2004, 26). While developed specifically to analyze crime as an alternative and/or compliment to hotspot mapping (Caplan and Kennedy 2011, 99–110) and other traditional analytical methods, RTM is also capable of analyzing political violence worldwide, given appropriately detailed data availability (Kennedy, Gaziarifoglu, and Caplan 2012).

RTM began in response to the need of state and local police to curb violent crime in the small township of Irvington, New Jersey, by forecasting where future events (particularly shootings) would occur (Caplan and Kennedy 2011, 15–16). A number of factors—drug, gang activity/presence, and infrastructure—were taken into account; individually, these factors correlated with the presence of shootings and when mapped separately and then combined into a composite map the factors suggested that "certain qualities of space coincide with the locations of shooting incidents" (Figure 1). From this composite map, analysts were able to forecast the probable locations of future shootings for a 6-month period through an examination of the last 6-month period. This process provided police a metric by which they could additionally measure the effectiveness of operations (the "treatment" effect) in the area of regard.

Figure 1. Composite map (right) from data layers (left) to forecast future shooting locations (Figure 2-4 in Caplan and Kennedy 2011, 17).
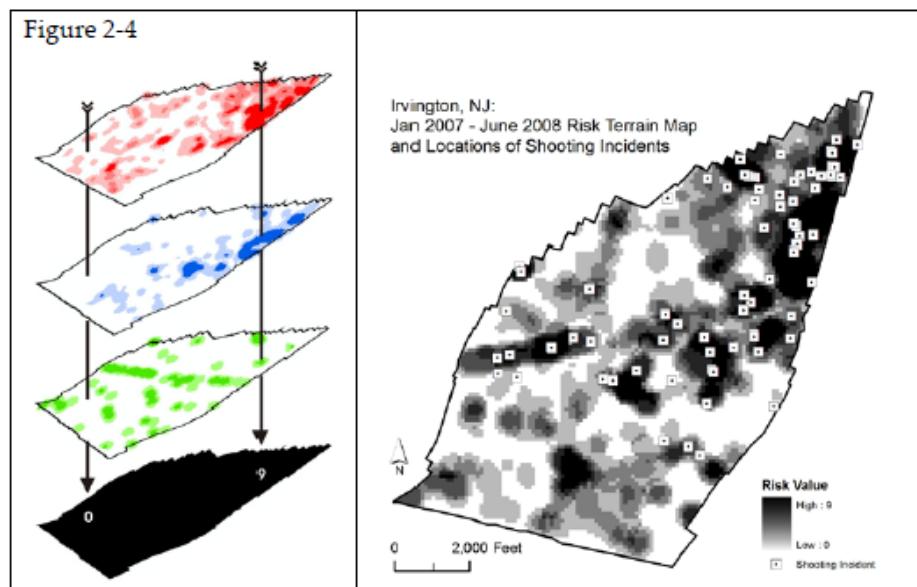
Data layers, however, were not chosen at random. A rather simple design was used to operationalize the data by gathering those data already collected, updated, and validated by the police. A density map was created using the points of gang members' residences, infrastructure (specifically the presence of strip clubs, bars, check cashing outlets, bus stops, pawn shops, fast-food restaurants, and liquor stores), and drug arrests (Caplan and Kennedy 2011, 18–19). The unit of analysis, rather than being the event itself in RTM, becomes the physical geography, terrain, and attributes associated with the area of regard—this suggests an analysis more focused on the complex interdependencies of systemic effects, rather than any one association with an event occurring.

Political event data have been used to understand international relations for years (King 1986, 1991; Gurr 1972; Alker 1975; Hilton 1976; Papayanopoulos 1973; Rai and Blydenburgh 1973; Rice 1926). While the great majority of such quantitative analyses have examined dyadic relationships between states and, more recently, conflict within states, some have begun to utilize highly localized data to explore specific issues, such as Lyall's (2009) examination of indiscriminate Russian artillery shelling of Chechen villages in the early 2000s to determine the impact of such shelling on insurgent attacks (see also O'Loughlin and Witmer 2011; Rustad et al. 2011). A key differentiation between these examinations and what RTM suggests, however, are the explanatory versus forecasting/risk analysis focus of the examinations. For instance, Lyall's 2009 study specifically sets out to test whether indiscriminate artillery strikes on villages have a positive effect on violence (positive in the sense that it increases or incites insurgent activity and violence). The RTM approach would use, for example, the prevalence of past violence, and the locations of known insurgent strongholds, local infrastructure, and previous artillery strikes to develop a systems approach to understanding and forecasting risk.

Likewise, the area of forecasting in the political science literature has seen great emphasis in the last few decades. Famine, humanitarian emergencies, tensions between groups, and natural disasters have become areas of great interest to governments, intergovernmental organizations, and nongovernmental organizations (on, the internet, users may find websites such as FEWSNET, FEWER, Ushahidi, FAST, and Crisis Watch for examples of early warning and emergency response systems). As noted in Kennedy, Gaziarifoglu, and Caplan (2012, 24), however, "a difficulty that

occurs with these is that they are often not very dynamic or complete," going on to say the "predictions that are made are often out of context and involve very 'linear' explanations." In order to compensate for the incomplete and static nature of these warning systems, Toomey and Kennedy (2011, 11–12) state that RTM helps to

> "solve certain resourcing issues, due to the lack of expensive specialist software/hardware required for it to function; enabling early warning systems to generate easily accessed and easily understood warnings through the use of GIS maps; improving risk assessment capabilities by increasing flexibility and facilitating integrated threat analyses, and by allowing for the inclusion of various different correlates and sources of information; and most importantly, explaining not only what threats are likely to occur in a certain area, but also to elaborate on the differential vulnerabilities of people within the area being studied."

One of the principal problems plaguing widespread analyses is the lack of an efficient means of producing highly precise geo-located event data from news articles, whether machine or human-coded. Once these data can be refined, however, and added to aggravating and mitigating risk factors, the possibilities of highly localizable RTMs to produce meaningful hazard-terrain surfaces for forecasting event likelihood are many—even in a global context. Data, it seems, will be the most limiting factor to analysis.[11]

Given the geographic specificity of event data with sufficient detail (neighborhood-level to facility-level geocoding) and the mission to provide forecasts for potentially politically-salient events tailored to said geographic specificity, RTM is a viable option for providing relevant data to support analysis and the intelligence preparation of the battlefield (IPB) process. Moreover, given Army analysts' penchant for analyzing a specific geographic area, RTM is ideal as, "forecasts based on risk assessments using RTM focus on the conditions of the environment where an event could occur. The unit of analysis is the geography, not the event." (Kennedy, Gaziarifoglu, and Caplan 2012, 16). The same authors make the case for utilizing RTM within the global context, specifying political event data to be used in lieu of police and law enforcement data to achieve similar results with risk assessment forecasts.

---

[11] For a comprehensive explanation and quick-start guide of the RTM methodology in a globalized/international context, see: Kennedy, Gaziarifoglu and Caplan, 2012, Ch. 2-5.

3.4 Dominant political narratives and event data

The VESCA work unit conducted a preliminary study to assess the feasibility of tracking dominant political narratives that may be found in media reports. The work is described in an ERDC-CERL special report (FOUO);[12] based on the approach used in the study, results suggest that it is feasible to achieve reasonably accurate classification of articles as representative of a narrative with machine automation. However, the approach used in the preliminary study was labor intensive. Automated detection of the presence of dominant political narratives in media reports offers the potential benefit of understanding events as related to "inform and influence" military activities. That term is described in ADRP 3-0 (2016) as "Inform and influence activities is the integration of designated information-related capabilities in order to synchronize themes, messages, and actions with operations to inform United States and global audiences, influence foreign audiences, and affect adversary and enemy decision making." Including dominant political narratives in event analyses can offer better understanding of strategic and operational considerations that may influence tactical activities, and the events that both frame the context for influence and explain how events may be interpreted by populations.

3.5 Situational understanding at tactical spatial mission scale

Existing geospatial analysis tools for small units (e.g., Squad, Special Operations Forces (SOF)) are not built for the order of complexity demanded by UO. Urban features contributing to analysis of maneuver, avenues of approach, fires, hazards, and communication are not well-defined and typically, these features are not collected, extracted, or stored in a manner to facilitate tactical analysis on a mission-ready device. In support of the ERDC project GATE, the VESCA team was asked to provide a potential subset of METT-TC, the set of mission variables described previously (Section 1.1). The VESCA team was specifically asked to provide civil considerations input to include in a GATE routing tool. Through collaboration with the GATE program team and Army subject matter experts, VESCA identified areas within the urban environment with potential symbolic meaning to the local population and areas that might increase risk of violence. VESCA team members prepared the data using political and violent events records from the datasets described in Chapter 2 and then, they enhanced

---

[12] ERDC SR-16-3; Distribution authorized to U.S. Government agencies: Administrative or Operational Use (11 July 2016). Other requests for this document shall be referred to Chief, CEERD-CNC of ERDC/CERL.

the event location data through a manual geocoding process. The resulting VESCA map layer was used to inform the GATE routing algorithm to influence the optimal route and provide a situational awareness overlay.

4 Enhancing Event Data

The VESCA work unit focused on three lines of effort related to event data analytics, including the following: location detail enhancement through geoparsing; data integration and deduplication, also referred to as data harmonization; and modeling and analysis for military application. This chapter reviews each of these lines of effort.

4.1 Geoparsing

4.1.1 Geoparsing background

A great deal of knowledge is available, in the form of unstructured text, from sources such as news reports and online content. In addition to processing the content of such sources, location information encoded into the text may provide additional opportunities for analysis and presentation. Print media articles commonly contain geographic metadata in the form of a dateline, which commonly includes the city-level location for the story (Zelizer and Allan 2010). Additional geospatial data must be extracted by using natural language processing (NLP), specifically through the subfield of named entity recognition (NER).

Natural language processing incorporates various machine learning and statistical techniques in order to process written text into a format understandable by computers. NER classifies text into categories of interest to the study, such as words likely to refer to a person, place, or specific part of speech. Following NER, a component known as a resolver is used to associate a placename with a record in a placename database (i.e., gazetteer), in order to retrieve the appropriate location and attributes (e.g., latitude and longitude). More background and implementation details may be found in Garfinkle et al. (2017).

Additional geoparsing implementation approaches and current geoparsing challenges are described in many recent publications; Leetaru (2012) and Lee, Liu, and Ward (2016) are both useful. Leetaru (2012) describes implementation approaches, toolkits, and commercial capabilities, along with details regarding the complexities of massive bulk processing and matching against gazetteers. Lee, Liu, and Ward (2016) delve into detailed discussion of identifying event-relevant locations and local place-name matching, especially on translated text and foreign language place-names.

Appendix C of this report describes some fundamental challenges for event models that also relate to geoparsing challenges.

The software developed for this project serves as a platform for experimenting on potential improvements to NER algorithms and toolsets, as applied to text news media sources. Geoparsing activities were limited to adaptation and evaluation of existing technologies, with follow-on steps exploring the ability to swap individual components.

The platform has been specified to enable the following four steps:

1. Input of plain text document.
2. NER through swappable natural language processing packages.
3. Resolve extracted entities by using a resolver and gazetteer.
4. Incorporate metrics of accuracy and confidence in order to compare results.

4.1.2   Stakeholders

This prototype was designed and implemented for users such that it can be utilized with minimal programming knowledge upon databases of plain-text print media articles such as ICEWS. Functionally, the requirement is specified so that the geoparsing tool can extract as many longitude and latitude coordinates as possible from poorly structured plain text content. Nonfunctionally, the software is structured so that users can compare different geoparsing technologies.

4.1.3   Operating environment

Per stakeholder consultations, the geoparsing capability of the prototype is designed to run on personal computers running U.S. Army Gold Standard Windows 7 and connected to a Department of Defense (DoD) NIPRNet (Nonclassified Internet Protocol Router Network). Because some interchangeable components require local administrative rights, the software is developed on the Engineer Research and Development Center (ERDC) Research and Development Engineering (RDE) network.

4.1.4   Existing components evaluated

To efficiently implement an effective geoparsing capability, commercial off-the-shelf (COTS) capabilities were acquired and examined.

Table 3. Existing natural language processing capabilities examined.

Language Name	Source(s)	Capabilities overview
CLAVIN and CLAVIN-NERD	https://github.com/Berico-Technologies https://clavin.bericotechnologies.com/	Java programs compiled using Maven and run as a virtual server. CLAVIN provides a number of features designed to help resolve ambiguous, misspelled, or alternatively-named place names. CLAVIN-NERD substitutes Stanford NLP.
CLIFF and CLIFF-UP	https://github.com/mitmedialab/CLIFF https://github.com/ahalterman/CLIFF-up	Server implementations of CLAVIN.
Mordecai	https://github.com/openeventdata/mordecai	Run as a web service and provides substantial control over search regions and for the substitution of custom NER models. Requires Docker. Built upon the MITIE information extraction library (https://github.com/mit-nlp/MITIE).
'Stock' Stanford CoreNLP	http://stanfordnlp.github.io/CoreNLP/	Utilized part of speech tagging to exhaustively search through all likely nouns as potential place names through brute-force resolving.
Custom-Trained Neural Network	https://keras.io/	Custom neural network implemented in Python using the Keras library.

4.1.5 Geoparsing implementation

Garfinkle et al. (2017) describes the initial implementation of geoparsing capabilities to satisfy VESCA project requirements. Figure 2 depicts an overview of the workflow and architecture implemented thus far.

Figure 2. Overview of geoparsing workflow and architecture.
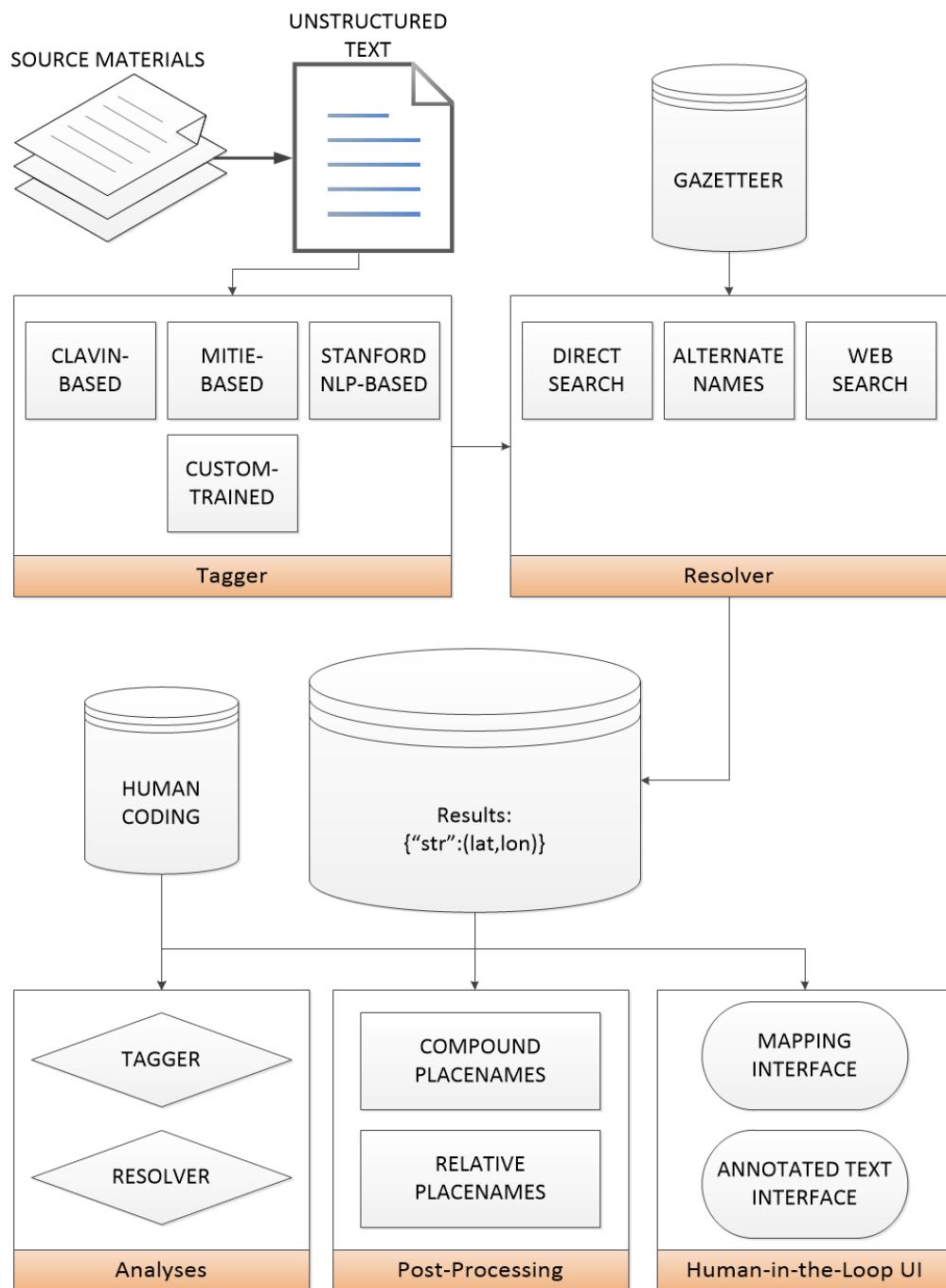
4.1.5.1 Geoparsing metrics

Goldberg (2008) defines metrics relevant to geocoding from a consoli-
dated set of messy address data, but the work does not define metrics for

when data begins with plain text that may or may not be related to place, and few or no actual address records.

Fundamental metrics referred to in information retrieval include precision and recall. Coppin (2004, 598) provides a useful definition of precision:

> *"If a system has 100% precision, it means that when it says that particular document is relevant, then it is guaranteed to be correct… Lower precision means that it will wrongly classify some documents as being relevant (false positives)."*

Coppin (2004, 598) also provides useful definition for recall:

> *"For a system to have 100% recall, it must be guaranteed to find all relevant documents within a corpus in response to a particular query. Lower recall means that the system will fail to identify some documents as being relevant (false negatives)."*

For geoparsing, the metrics relate to the NER, the resolver, and the gazetteer. Figure 3 depicts relationships between geoparsing and false positives and negatives.

Figure 3. Geoparsing sources of false positives and false negatives (ERDC-CERL).
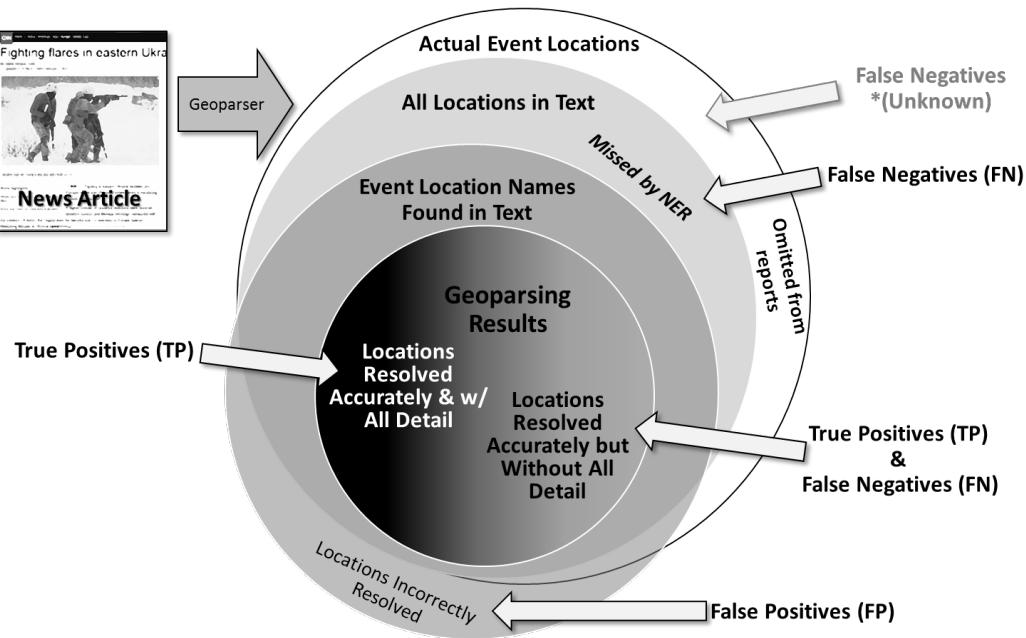
In order to test and evaluate components of the flexible architecture depicted in Figure 2, evaluation must separately consider precision and recall for the tagger, and precision and recall of the resolver. Additionally, the gazetteer presents additional challenges that affect resolver performance.

4.1.5.2 Dataset preparation

NOTE: Portions of this subsection are not included in this unclassified publication; content removed has been noted. See Volume 2, the limited distribution version of this publication for FOUO content.

The VESCA team acquired ICEWS data and enhanced it with human-in-the-loop coding. The team queried the dataset for event types of "Demonstrate or rally" and "Protest violently, riot" for three cities over multiple time periods, and downloaded the events and articles yielded by the ICEWS platform. Query results are summarized in Table 4.

Table 4. (*FOUO content removed, including figure.*)

Through a human-in-the-loop process, the VESCA team extracted additional event information. Coders identified placenames indicating the location of events and the spatial resolution of the coded location (Named Populated Place/City; Named District/Neighborhood in Populated Place – Upazila, Commune, Subdivisions/Raions; Named Spot/Building/Area; or Roadway). The team also recorded when an article described events as occurring at multiple locations, and if the dataset included duplicate or erroneous event entries. When additional detailed location information was not included in the ICEWS article, other online articles were found and acquired that did offer additional detailed location information.

(FOUO content removed here.)

Following initial assessment of feasibility, senior team members reviewed the dataset and articles to prepare a gold-standard placename dataset for a set of events and articles. The dataset uses placename information solely available from information contained in the article. This dataset includes the complete original article text, a corresponding list of placenames referenced in the text, and identification of placenames that are the locations of the described event. The placenames are associated with the results of

matching places in several gazetteers, including Geonames, Open Street Map, Wikimapia, and Google Places. For each match, the entity's identification (ID) and other entity characteristics are recorded to correspond with the event. By preparing the gold-standard dataset using this method, personnel may use article text and identified placenames to test NER capabilities and resolver functions. As geoparsing implementation proceeds in later work, this dataset may be used to evaluate component performance and overall performance.

4.2   Data harmonization

Armstrong et al. (2015) describes the adaptation of the Actionable Intelligence Retrieval System (AIRS) to support event data harmonization – the alignment of multiple datasets to a common schema, and the tools and process to identify duplicate entries. The preliminary capability described in the 2015 paper used the Karma[13] user interface to prepare data models to align the ACLED, SCAD, UCDP, GED, and UCDP actors datasets. It also included an event resolution scoring model (McConky 2012) to enable an analyst to determine event co-references (i.e., likelihood that two data records refer to the same event).

Continued development since 2015 to support VESCA has extended the data harmonization prototype. Detailed description and instructions about the capability are presented in Appendix D. In summary, the capability now includes improved data ingestion workflows, and ready-built data models to accommodate data from sources such as those identified in section 2.2, including GTD, ACLED, SCAD, UCDP, ICEWS, and the PDP. The prototype has been extended to support optional extraction of additional entity information using natural language processing of article text. Additionally, the capability now allows users to develop, maintain, and select among multiple scoring models for event data resolution to support the user in determining whether two events are likely to refer to the same actual event instance. Lastly, the new version supports data export to common data formats for ingest into other analytical tools and platforms.

4.3   Military modeling and analysis example

This section builds off prior sections to summarize an example process for event data enhancement, fusion, and analytics to yield mission-relevant

---

[13] http://www.isi.edu/integration/karma/

information. Chapter 3 describes mission relevance of political event data
– presenting potential opportunities for why and how event data may be
used. Earlier sections in Chapter 4 describe capabilities emerging from
VESCA to support enhancing geographic details and harmonizing event
datasets. Appendix C provides detailed discussion of additional event
modeling possibilities.

The process developed under VESCA to transform event data from diverse
sources into a mission-relevant form involves several steps, as summa-
rized in Figure 4.

Figure 4. Process to transform event data into mission-relevant information.
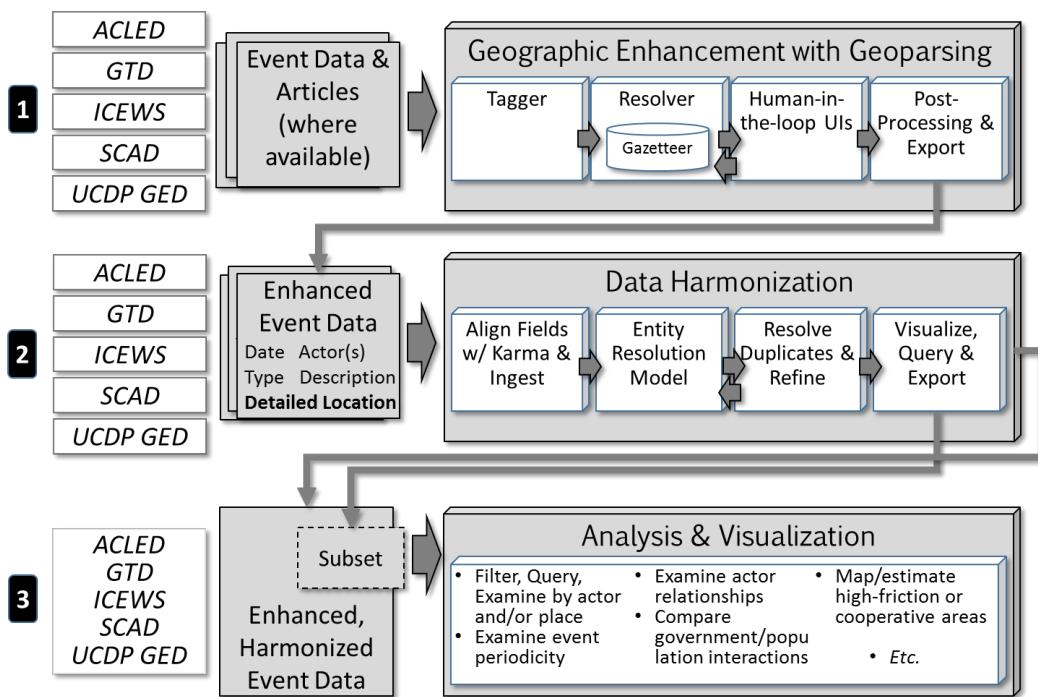
Row 1 of Figure 4 depicts the following steps and data:

    A. Acquire datasets, such as:
        a. ACLED
        b. GTD
        c. ICEWS (filtered to events in Bangladesh during limited time periods)
        d. SCAD
        e. UCDP GED
    B. Enhance geographic details through geoparsing and human-in-the-loop interaction

a. Tag location mentions.
b. Resolve location mentions to geographic features.
c. Train and validate resolved locations with human-in-the-loop user interfaces.
d. Export location-enhanced dataset(s).

Row 2 of Figure 4 depicts the following steps:

A. Prepare harmonization tool for dataset ingest by loading or configuring Karma data models. Additional Karma information may be found in Appendix D and at http://usc-isi-i2.github.io/karma/.
   a. Ingest datasets into data harmonization tool.
   b. Select subset for entity resolution (if needed), such as a country (e.g., Bangladesh).
B. Prepare and execute entity resolution scoring model (or reuse existing preconfigured scoring models) to detect duplicate event entries.
   a. Review entity resolution results (i.e., entries deemed duplicates); determine whether scoring model is appropriate or identifying duplicate entries incorrectly (e.g., ICEWS events that are at multiple locations in a city being deemed duplicates).
C. Edit and refine scoring model until duplicate entries are identified, while minimizing removal of legitimately separate event entries.
D. Visualize results in harmonization tool, execute query, and export results as a comma-separated-value (CSV) table, or directly ingest into analytic tool.
   a. Select bounding box or enter search for placename (e.g., Dhaka) and export.

The process depicted in Row 3 of Figure 4 begins with event data harmonized to a common schema, with duplicates excluded, and includes the following steps:

A. Ingest harmonized and enhanced data into analytic tool, such as Esri ArcGIS or ERDC's STRIDER application.
B. Interact with dataset to support analyses described in Chapter 3. Such analyses may include:
   a. Query the event dataset for a country, city, or neighborhood to extract named groups of interests relevant to the area, the

types of events they are involved in, the groups to which they are oppositional or cooperative, and potential motivating factors precipitating violence or other engagement.

b. Identify where a group has or a collection of groups have previously focused their activities.

c. Examine historical event data to determine group disposition for oppositional activities against groups, national government, NGOs, IGOs, and U.S. interests. Conversely, examine group disposition for supportive activities toward groups, national government, NGOs, IGOs, and U.S. interests.

d. Establish a baseline for political violence in the area – how many cooperative events are recorded over preceding time periods; how many confrontational events; how have event types changed over time and place?

e. Map actor or group activities with geospatially specific data to indicate where certain groups are most active, and with what types of events; does the event data and descriptive text suggest those areas are likely controlled by those groups, or are they merely where they are active?

f. Map event types with geospatially specific data to indicate where such activities have been common; would such activities be impacted or impact U.S. operations?

g. Summarize findings in graphic, tabular, and text form for stakeholders, such as:

   i. Ordered list of actors/groups and their relative cooperation/conflict with U.S. groups.

   ii. Ordered list of actors/groups and their relative cooperation/conflict with other significant stakeholders.

   iii. Series of heat maps of geographically enhanced event data showing density of reports of event types over space, organized by particular event types or Goldstein score.

   iv. Summary heat map showing collections of select event types.

   v. Map and list of locations where major violent or other significant events have previously occurred.

   vi. Heat map showing results of application of risk terrain modeling approach to event types of greatest interest – yielding, for example, potential locations of future protests, demonstrations and riots.

5 Summary and Recommendation

This report shows that the VESCA team developed and demonstrated pro-cesses and tools to enhance and harmonize violent event data collected from diverse sources, thereby enabling an analyst to evaluate and integrate multiple sources of data, work with enhanced event data spatial resolution, and analyze and/or visualize the data to produce mission-relevant infor-mation. The report includes background on event data sources; study of protests, demonstrations, and rallies; and relevant analytical methods. The report describes doctrine regarding civil considerations, sociocultural analysis, and contingency basing; those sections present how event data may be transformed from its original tabular or text format and inter-preted to support doctrinal analysis. The report also describes how event data may be enhanced through geoparsing and harmonization processes—to align datasets to a common schema and to identify duplicate event en-tries. Finally, the report describes how data may be analyzed and pro-cessed to yield mission-relevant results.

In concluding this portion of the work package, the VESCA team has demonstrated progress on event data harmonization by implementing and using a prototype to align event data sources to a common schema and to identify and resolve duplicate events. VESCA manually enhanced details of event locations to produce political event data that could be analyzed with greater spatial precision and was sufficiently detailed to be operationally and tactically mission-relevant. The report describes how VESCA work was incorporated into work package demonstrations for GATE, STRIDER and ENSITE. This report also describes progress on prototyping automa-tion of spatial enhancement (Appendix D). While automation of sufficient quality has not yet been achieved, VESCA confirms that many event arti-cles provide sufficient information to extract such spatial information.

It is recommended that future work should continue to improve processes for tagging foreign and translated placenames, while also resolving such placenames efficiently and effectively with reliable gazetteers. Relevant government and commercial technology development continues in this do-main and when such technology is mature, it could be used in place of the manual processing or prototype components described in this report.

References

ADRP 2-0. 2014. Intelligence Operations. Washington, DC: HQDA.

ADRP 3-0. 2016. *Operations*. Washington, DC: Headquarters, Department of the Army.

ADRP 3-07. 2012. *Stability*. Washington, DC: Headquarters, Department of the Army.

Al-Chaar, Ghassan, George W. Calfas, Michael A. Weiss, Michael K. Valentino, and
    Patrick J. Guertin. 2016. *Construction Material-Based Methodology for Military
    Contingency Base Construction: Case Study of Dhaka, Bangladesh*. ERDC-
    CERL TR-16-14. Champaign, IL: U.S. Army Engineer Research and Development
    Center-Construction Engineering Research Laboratory (ERDC-CERL).

Alker, Hayward R., Jr. 1975, "Polimetrics: Its Descriptive Foundations." In *Handbook of
    Political Science*, Fred Greenstein and Nelson Polsby, editors, 139–210. Reading,
    MA: Addison-Wesley.

Armstrong, Chandler, Ryan M. Brown, Jillian Chaves, Adam Czerniejewski, Justin Del
    Vecchio, Timothy K. Perkins, Ron Rudnicki, and Greg Tauer. 2015. "Next
    Generation Data Harmonization." In *Proceedings SPIE Conference* Vol. 9499,
    Next-Generation Analyst III, 94990D. International Society for Optics and
    Photonics, conference held April 20, 2015 in Baltimore, MD.
    doi:10.1117/12.2180458.

(ARCIC) Army Capabilities Integration Center. "Army Warfighting Challenges." Accessed
    9/8/2017. http://www.arcic.army.mil/Initiatives/armywarfightingchallenges.

ATP 2-01.3. 2014. "Intelligence Preparation of the Battlefield." Washington, DC:
    Headquarters, Department of the Army (HQDA)

ATP 3-37.10. 2017. "Base Camps." Washington, DC: HQDA.

Azar, Edward E. 1975. "The Issues of Event Research." In *Theory and Practice of Events
    Research: Studies in Inter-nation Actions and Interactions*, Vol. 1:91-119, E. E.
    Azar and J. D. Ben-Dak, editors. New York: Gordon and Breach Science
    Publishers.

_____. 1993. "Conflict and Peace Data Bank (COPDAB), 1948-1978 [Computer
    file]," 3rd release. College Park, MD: University of Maryland, Center for
    International Development and Conflict Management [producer]. Ann Arbor,
    MI: Inter-university Consortium for Political and Social Research (ICPSR)
    [distributor].

_____. 2009. "Conflict and Peace Data Bank (COPDAB), 1948–1978" [Computer
    file]. ICPSR07767-v4. College Park, MD: University of Maryland, Center for
    International Development and Conflict Management [producer]. Ann Arbor,
    MI: Inter-university Consortium for Political and Social Research (ICPSR)
    [distributor]. https://doi.org/10.3886/ICPSR07767.v4.

Beieler, John, P. T.Brandt, A. Halterman, P. A. Schrodt, and E. M. Simpson. 2016. "Generating Political Event Data in Near Real Time." In *Computational Social Science: Discovery and Prediction* (reprint edition), 98. Cambridge, England: Cambridge University Press.

Bond, Doug, Joe Bond, Churl Oh, J. Craig Jenkins, and Charles Lewis Taylor. 2003. "Integrated Data for Event Analysis (IDEA): An Event Typology for Automated Events Data Development." *Journal of Peace Research* 40 (6): 733–45.

Caplan, Joel M. (au-ed.), and Leslie W. Kennedy, ed. 2011. *Risk Terrain Modeling Compendium*. Newark, NJ: Rutgers Center on Public Security.

Chojnacki, Sven. 2012. "Event Data on Armed Conflict and Security: New Perspectives, Old Challenges, and Some Solutions." *International Interactions* 38(4): 382–401.

Coppin, Ben. 2004. *Artificial Intelligence Illuminated*. Burlington, MA: Jones & Bartlett Learning.

EP 1105-3-1. 2009. Base Camp Development in the Theater of Operations. Washington, DC: Headquarters, U.S. Army Corps of Engineers (HQUSACE).

FM 3-13. January 2013. Inform and Influence Activities. Washington, DC: Headquarters, Department of the Army (HQDA).

FM 3-57. 2014. Civil Affairs Operations. Washington, DC: HQDA.

Garfinkle, Noah W., Lucas Selig, Timothy K. Perkins, and George W. Calfas. 2017. "Geoparsing Text for Characterizing Urban Operational Environments Through Machine Learning Techniques." *SPIE Defense+ Security*, Conference Volume 10199. International Society for Optics and Photonics. doi: 10.1117/12.2262808.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "Machine Coding of Event Data Using Regional and International Sources." *International Studies Quarterly* 38(1) 91–119.

Gerner, Deborah J., Philip A. Schrodt, Omur Yilmaz, and Rajaa Abu-Jabr. 2002. "The Creation of Cameo (Conflict and Mediation Event Observations): An Event Data Framework for a Post Cold War World." Presented at the annual meeting of the American Political Science Association held Aug 28 at Boston, MA.

Goldberg, Daniel W. 2008. *A Geocoding Best Practices Guide*. Submitted to The North American Association of Central Cancer Registries (NAACCR) on November 10, 2008. Los Angeles, CA: University of Southern California, GIS Research Laboratory.

Goldstein, J. S. 1992. A Conflict-Cooperation Scale for WEIS Events Data. *Journal of Conflict Resolution* 36(2): 369-385.

Gupta, S., P. Szekely, C.A. Knoblock, A. Goel, M. Taheriyan, and M. Muslea. May 2012. "Karma: A System for Mapping Structured Sources into the Semantic Web." In *Proceedings of 9th Annual Extended Semantic Web Conference*, 430–434. Berlin, Heidelberg: Springer.

Gurr, Ted Robert. 1972. *Polimetrics: An Introduction to Quantitative Macropolitics*. Englewood Cliffs, NJ: Prentice Hall.

Hilton, Gordon. 1976. *Intermediate Polimetrics*. New York: Columbia University Press.

Kennedy, Leslie W., and Erin Gibbs Van Brunschot. 2009. *The Risk in Crime*. Lanham, MD: Roman and Littlefield.

Kennedy, Leslie W., Yasemin Gaziarifoglu, and Joel M. Caplan. 2012. *Analyzing and Visualizing Worldwide Spatial Data: An Application of Risk Terrain Modeling*. Newark, NJ: Rutgers Center on Public Security.

King, Gary. 1986. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30(3): 666–687.

_____. 1991. "On Political Methodology." *Political Analysis* 2: 1–30.

Lee, Sophie J., Howard Liu, and Michael D. Ward. 2016. "Lost in Space: Geolocation in Event Data." arXiv preprint available at arXiv:1611.04837.

Leetaru, Kalev. 2010. "The Scope of FBIS and BBC Open Source Media Coverage, 1979–2008." *Studies in Intelligence* 54(1): 51–71.

_____. 2012. "Fulltext Geocoding Versus Spatial Metadata for Large Text Archives: Towards a Geographically Enriched Wikipedia." *D-Lib Magazine* 18.9/10.

_____. 2015. "Why We Can't Just Read English Newspapers to Understand Terrorism: And How Big Data Can Help." *Foreign Policy*. http://foreignpolicy.com/2015/04/15/why-we-cant-just-read-english-newspapers-to-understand-terrorism-big-data/.

Lyall, Jason. 2009. "Does Indiscriminate Violence Incite Insurgent Attacks: Evidence from Chechnya." *Journal of Conflict Resolution* 53(3): 331–362.

McClelland, Charles A. 1978. "World Event/Interaction Survey Codebook." Third ICPRS Edition. Ann Arbor: Inter-University Consortium for Political and Social Research.

McConky, K.T. 2012. "Applications of Location Similarity Measures and Conceptual Spaces to Event Coreference and Classification." Dissertation. Buffalo, NY: State University of New York.

Merritt, Richard L., Robert G. Muncaster, and Dina A. Zinnes.1993. *International Event-Data Developments: DDIR Phase II*. Ann Arbor, MI: University of Michigan Press.

O'Brien, Sean P. 2010. "Crisis Early Warning and Decision Support: Contemporary Approaches and Thoughts on Future Research." *International Studies Review* 12(1): 87–104. doi: 10.1111/j.1468-2486.2009.00914.x.

O'Loughlin, John, and Frank D. Witmer. 2011. "The Localized Geographies of Violence in the North Caucasus of Russia, 1999–2007." *Annals of the Association of American Geographers* 101(1): 178–201.

Papavanopoulos, L. 1973. "Democratic Representation and Apportionment: Quantitative Methods, Measures, and Criteria." *Annals of the New York Academy of Sciences* 219: 3–4. doi: 10.1111/j.1749-6632.1973.tb41397.x.

Rai, Kul B., and John C. Blydenburth. 1973. *Political Science Statistics*. Boston: Holbrook Press.

Rice, Stuart A. 1926. "Some Applications of Statistical Method to Political Research." *American Political Science Review* 20(2): 313–329.

Rustad, Siri Ass, Halvard Buhaug, Ashild Falch, and Scott Gates. 2011. "All Conflict is Local: Modeling Subnational Variation in Civil Conflict Risk." *Conflict Management and Peace Science* 28(1): 15–40.

_____. 2001. "Automated Coding of International Event Data Using Sparse Parsing Techniques." Presented at annual meeting of the International Studies Association held 23 February in Chicago, IL.

_____. 2014. "TABARI - Textual Analysis by Augmented Replacement Instructions Version 0.8.4." Charlottesville, VA: Parus Analytical Systems.

_____. 2015. "Event Data in Forecasting Models: Where Does It Come From, What Can It Do?" Presented at Conference on Forecasting and Early Warning of Conflict, Peace Research Institute, held in Oslo (Norway), 22–23 April.

Schrodt, Philip A., Ömür Yilmaz, Deborah J. Gerner, and Dennis Hermrick. 2008. "Coding Sub-State Actors using the CAMEO (Conflict and Mediation Event Observations) Actor Coding Framework." In *Annual Meeting of the International Studies Association*, held 26–29 March in San Francisco, CA.

Schrodt, Philip A., John Beieler, and Muhammed Iris. 2014. "Three's a Charm?: Open Event Data Coding with EL:DIABLO, PETRARCH, and the Open Event Data Alliance." Presented at International Studies Association meeting, March 26-29, Toronto.

Shellman, S. M. 2008. "Coding Disaggregated Intrastate Conflict: Machine Processing the Behavior of Substate Actors over Time and Space." *Political Analysis* 16(4): 464–477.

Toomey, M., and Leslie W. Kennedy. 2011. "An Analysis of Modern Early Warning Systems: How Might Risk Terrain Modeling Contribute to the Development of an Optimal System?" Newark, NJ: Rutgers Center on Public Security.

TRADOC (U.S. Army Training and Doctrine Command). 2009. The United States Army Concept Capability Plan for Army Base Camps in Full Spectrum Operations for the Future Modular Force 2015–2024. TRADOC Pamphlet 525-7-7. Fort Monroe, VA: Headquarters, TRADOC.

Tuchinda, Rattapoom, Craig A. Knoblock, and Pedro Szekely. 2011. "Building Mashups by Demonstration." *ACM Transactions on the Web (TWEB)* 5(3): 1–50.

UN/ISDR (United Nations Inter-Agency Secretariat of the International Strategy for Disaster Reduction). 2004. *Living with Risk* (Vol. 2). Geneva: UN Publications.

U.S. Army. 2008. *Political Military Analysis Handbook*. Version 3.3. Fort Bragg, NC: John F. Kennedy Special Warfare Center and School.

USCENTCOM (U.S. Central Command). 2009. Construction and Base Camp Development in the USCENTCOM Area of Responsibility (commonly known as "The Sand Book"). Regulation 415-1. Tampa, FL: MacDill Air Force Base, Headquarters USCENTCOM.

USFK (U.S. Forces, Korea). 2004. "Host Nation Funded Construction in Korea." USFK Regulation 415-1. Seoul, Republic of Korea: Headquarters, USFK.

Veen, Tim. 2008. "Event Data: A Method for Analysing Political Behaviour in the EU." Paper delivered at the Fourth Pan-European Conference on EU Politics, held 25–27 September 2008 at Riga, Latvia. Nottingham, UK: University of Nottingham.

Ward, Michael D., Andreas Berger, Josh Cutler, Matthew Dickenson, Cassy Dorff, and Benjamin J. Radford. 2013. "Comparing GDELT and ICEWS EVENT Data." White Paper. Durham, NC: Duke University. https://www.researchgate.net/publication/303211430_Comparing_GDELT_and_ICEWS_event_data.

Weidmann, Nils B. 2016. "A Closer Look at Reporting Bias in Conflict Event Data." *American Journal of Political Science* 60(1): 206–218.

Yonamine, J. E. 2013. "A Nuanced Study of Political Conflict Using the Global Datasets of Events Location and Tone (GDELT) Dataset." Doctoral dissertation, The Pennsylvania State University. https://etda.libraries.psu.edu/catalog/18659.

Appendix A: Excerpts of Army Documents

Excerpts from Army documents are provided below that are relevant to Section 3.1, "The importance of situational understanding for contingency base site selection." Note that the numbers given at the beginning of each item represent the paragraph number, as used within the document.

NOTE: Portions of this appendix are not included in this unclassified publication; content removed has been noted. See Volume 2, the limited distribution version of this publication, for FOUO content.

Base Camps (ATP 3-37.10)

"1-88. The intelligence section serves as the principal staff for providing intelligence to support current and future operations and plans. This section gathers and analyzes information on enemy, terrain, weather, and civil considerations for the base camp commander/BOS-I."

"1-94. The G-9/S-9 advises the base camp commander/BOS-I on the military operations effect on civilians in the AO relative to the complex relationship of civilians with the terrain and institutions over time. The G-9/S-9 is responsible for enhancing the relationship between Army forces, the civil authorities, and people in the AO."

"2-46. The goal of base camp site selection is finding the best possible location for a base camp that balances mission, sustainment/CSS, protection/force protection, environmental considerations, and construction requirements. Site selection, the actual process of choosing a site, occurs later in the base camp planning process. Selecting the best location for a base camp is a balance between operational, sustainment, and construction requirements. It also involves consideration of the operational and mission variables. The selection of a base camp site occurs after the preliminary planning phase."

"2-60. Base camp planning identifies when, where, and why base camps are needed and the details of life cycle activities. Base camp planning begins as part of crisis action planning, is part of a campaign and major operation planning, and continues through OPLAN and OPORD development and execution."

"2-64. Base camp planning requires a combined arms approach to harness the necessary expertise in the fields of sustainment/logistics, engineering, AT, protection, civil affairs, environmental resources, PVNTMED, resource management, safety, law, ranges and training areas, contracting, real estate, as well as other fields. It involves the unit staff of the primary organization that will be occupying the base camp, higher headquarters, and representatives from supporting units and organizations."

Table B-1[14] (B-3 – B-4)

"B-8. The staff determines possible locations for base camps based on an analysis of operational and mission variables, with added emphasis on terrain, civil, and environmental considerations."

Table B-2[14] (B-5)

"B-9. Site selection begins during mission analysis/problem framing with the identification of suitable and unsuitable areas that aims to narrow options and facilitate timely COA development. ...Unsuitable areas, which should generally be avoided, include areas such as those that are prone to flooding, have severe slopes or dense vegetation, or are inaccessible to heavy construction equipment and areas that are environmentally sensitive or that have historical, cultural, or religious significance."

"B-32. Base camp information requirements are identified collectively, and then selected staff members gather the necessary information within their area of expertise through their respective staff section or through reachback. ...[bullet] Local government and population attitudes on base camps and/or willingness to cooperate and provide assistance."

(FOUO Content Removed – this subsection)

(FOUO content removed here.)

(FOUO content removed here.)

---

[14] Tables are reproduced at end of this appendix.

Intelligence ADRP 2-0

"5-15. ASCOPE characteristics (areas, structures, capabilities, organizations, people, and events) are used to analyze and describe civil considerations that may affect operations. Included in civil considerations analysis are the effects urban centers may have on friendly and threat forces. There is no standard product resulting from this analysis. The G-2/S-2 generally develops products that fit the information needed to describe the situation and support thee commander's situational understanding." (page 5-3)

(FOUO Content Removed – this subsection)

(FOUO content removed here.)

Intelligence Preparation of the Battlefield ATP 2-01.3

"3-18. Civil considerations reflect the influence of manmade infrastructure, civilian institutions, and attitudes and activities of the civilian leaders, populations, and organizations within the operational environment on the conduct of military operations. Commanders and staffs analyze civil considerations in terms of the categories expressed in the memory aid ASCOPE (areas, structures, capabilities, organizations, people, and events)."

"3-19. Civil considerations help commanders understand the social, political, and cultural variables within the AO and their effect on the mission. Understanding the relationship between military operations and civilians, culture, and society is critical to conducting operations and is essential in developing effective plans. Operations often involve stabilizing the situation[,] securing the peace, building host-nation capacity, and transitioning authority to civilian control. Combat operations/*major operations* directly affect the populace, infrastructure, and the force's ability to transition to host-nation authority. The degree to which the populace is expected to support or resist U.S. and friendly forces also affects offensive and defensive operational design."

"3-20. ... Commanders consider how [social, economic, and political] factors may relate to potential lawlessness, subversion, or insurgency. Their goal is to develop their understanding to the level of cultural awareness."

"3-21. To improve the commanders sociocultural understanding, intelligence staffs can use sociocultural databases and repositories as well as HTTs/*foreign area officers, regional affairs officers, and other cultural enablers*, when available, to aid in the intelligence analysis conducted as part of assessing civil considerations."

"4-114. Events are routine, cyclical, planned, or spontaneous activities that significantly affect organizations, people, and military operations."

"10.32. Template or analyze faction activity as it relates to past events to analyze potential trends."

"A-25. As there are many different categories of civilians, there are many categories of civilian events that may affect the military mission. Some examples are planting and harvesting seasons, elections, riots, and voluntary and involuntary evacuations. Likewise, there are military events that impact the lives of civilians in an AO. Some examples are combat operations, including indirect fires, deployments, and redeployments. Civil-military operations planners determine what events are occurring and analyze the events for their political, economic, psychological, environmental, and legal implications." (A-14)

(FOUO Content Removed – this subsection)

(FOUO content removed here.)

Stability ADRP 3-07

"1-10. Addressing the drivers of violent conflict begins with a thorough assessment. The assessment analyzes the conditions of an operational environment, including how the operations affects the situation on the ground and how locals perceive the conditions."

Civil Affairs Operations FM 3-57

"1-15. During the military decision-making process (MDMP), CA Soldiers on the CAO staff (G-9/S-9) provide the commander with an analysis of the civil components that shape the operational environment. …The CAO staff provides the commander detailed civil considerations analysis focused on the factors (areas, structures, capabilities, organizations, people, and events [ASCOPE]) affecting the civil component of the AO."

"1-22. [bullet] Developing an analysis using ASCOPE to determine—

What, when, where, and why personnel might encounter civilians in the AO.

What activities civilians in the AO are engaging in that might affect the military operation (and vice-versa).

What the commander must do to support or interact with civil actions."

"3-43. Civil information is information developed from data with relation to civil areas, structures, capabilities, organizations, people, and events within the civil component of the commander's operational environment."

"3-46. Collection is the first step of the CIM process and refers to the literal gathering of relevant data. Driven by the CCIR and integrated with the ISR plan, civil information collection occurs at all levels through CR, data mining and collaboration with IPI, IGOs, NGOs, and OGAs. At first there is little, if any, quality screening of the data collection, everything related is relevant."

"3-47. About 90 percent of intelligence starts as open-source information."

Tables from the above doctrines

Base Camps ATP 3-37.10

Table B-1. Base camp planning considerations during the planning process

Steps of the MDMP	Steps of the MCPP	Base Camp Planning Considerations
Receipt of the mission	Problem framing	• Identify potential sources of data and information to include existing assessment products such as EBSs, OEHSA, and infrastructure assessments. • Request geospatial information and terrain visualization products to help understand terrain effects. • Request intelligence products on potential threats to the base camp. • Gather information on the local population to determine its effect on possible base camp locations. • Update running estimates/staff estimates. • Disseminate base camp-relevant information to the appropriate staff sections for inclusion in their running estimates/staff estimates.

Mission analysis	Problem framing (continued)	• Understand the higher command's basing strategy.
		• Assess assets available to perform base camp life cycle activities (joint and multinational forces, host nation, and contractors), identify obvious shortfalls, and prepare requests for augmentation for the commander's approval.
		• Determine constraints to include—
		  ▪ Allowable design and construction standards in theater-specific guidelines.
		  ▪ Higher headquarters policies, procedures, plans, orders, and directives.
		  ▪ Joint and Army/Marine Corps directives and regulations.
		  ▪ International and U.S. laws and regulations as applicable.
		  ▪ Host-nation laws and local customs and practices.
		• As part of the initial intelligence preparation of the battlefield/battlespace—
		  ▪ Evaluate terrain and weather effects on base camp activities.
		  ▪ Evaluate the effects of adversaries and neutrals on base camp activities.
		  ▪ Assess the availability of existing facilities and infrastructure within the operational area, and develop facts and assumptions to support assessments.
		  ▪ Identify potential base camp locations based on threat patterns and terrain.
		• Identify specified and implied base camp tasks and recommended essential base camp tasks, determine any obvious shortfalls in assets available, and initiate requests for support or augmentation as early during planning as possible.
		• Integrate information requirements and engineer or other necessary specialized reconnaissance capabilities into the information collection plan.

Table B-1. Base camp planning considerations during the planning process (continued)

Steps of the MDMP	Steps of the MCPP	Base Camp Planning Considerations

COA development	COA development	• Integrate the base camp principles.
		• Refine base camp requirements and possible solutions based on mission variables.
		• Recommend base camp locations based on the—
		  • Availability of existing facilities and infrastructure.
		  • Terrain, environmental, and civil considerations.
		  • Threats to base camps.
		  • Ability to sustain and secure base camps in a specific area.
		• Allocate base camp capabilities based on identified requirements (troop-to-task analysis).
		• Identify nodes and linkages of base camps, including the formation of base clusters.

COA analysis	COA wargaming	• Identify advantages and disadvantages of base camp design solutions using the following evaluation criteria developed before wargaming, such as—
		  • **Protect.** The ability to employ response forces and first responders in response to attacks and emergencies.
		  • **Sustain.** The ability to access base camps for services, resupply, and casualty evacuation.
		  • **Maneuver.** Mission support to maneuver units
		• Wargame (action/reaction) enemy attacks and emergencies on base camps and the employ response forces and first responders.

COA comparison	COA comparison and decision	• Analyze and evaluate advantages and disadvantages of each COA from a base camp perspective using the evaluation criteria developed before wargaming.
COA approval		• Gain approval for any changes to the essential tasks for base camps. • Gain approval for recommended priorities of effort and support. • Gain approval for requests for base camp augmentation to be sent to higher headquarters. • Initiate real estate acquisition actions once base camp locations have been approved. • Provide commander with updates on base camp issues or concerns within the COA decision briefing as appropriate.
Orders production, dissemination, and transition	Orders development Transition	• Integrate base camp tasks within the plan or order, and produce the base camp appendix. • Ensure the quality and completeness of the subordinate unit's instructions for performing base camp life cycle tasks.

Note. The Army uses the MDMP, and the Marine Corps uses the MCPP. The processes are similar, although the steps are different. The MDMP is described in FM 6-0; the MCPP is described in MCWP 5-10.

Legend:

COA	course of action
EBS	environmental baseline survey
FM	field manual
MCPP	Marine Corps planning process
MCWP	Marine Corps warfighting publication
MDMP	military decision-making process
OEHSA	occupational and environmental health site assessment
U.S.	United States

Table B-2. Site selection considerations in relation to mission variables (METT-TC/METT-T)

Mission Variables	Site Selection Considerations

Mission	• Analyze the unit mission to determine the purpose of base camps and the major functions they must perform based on tenant and transient unit operational requirements, to include—
	  ▪ Requirements for specific types of facilities such as airfields, landing zones, ammunition supply points, and firing ranges.
	  ▪ Types and sizes of tenant units (land area requirements).
	  ▪ Future requirements (sufficient land area for expansion; accessibility; and access to sources of water, power, and energy).

Enemy	• Analyze threats to the base camp and the associated protection considerations such as proximity to populations, standoff, and perimeter requirements.
OAKOC/KOCOA	• Vegetation—effects on movement, landing zones, observation, and cover and concealment. • Hydrology—access to water and avoidance of surface drainage. • Soil composition—suitable for construction, trafficability, and waste management options. • Surface and subsurface configuration—trafficability; cut, fill, and clearing requirements; natural slope for drainage; seismic conditions; and clear line-of-sight for communication and collection systems. • Obstacles—natural and man-made impediments (including the presence of people) to base camp construction, operations, and sustainment. • Man-made features—existing structures and local facilities and infrastructure that affect base camps.
Troops and support available	• Availability of local workers, equipment, and services to perform base camp construction and operations tasks.
Time available	• Time available for construction (based on when the constructing unit can occupy the site and the delivery of construction materials and equipment).

ASCOPE	• Relationship with the local population (acceptance and tolerance).
	• Local political climate and perceptions and the effects on location, design, and land use decisions. Politically unpopular decisions may attract acts of aggression.
	• Effects traffic, explosive safety, inconvenience on adjacent landowners.
	• Proximity to historical, cultural, religious, and environmentally sensitive areas.
	• Areas to include—
	  ▪ Sources of natural construction resources (water, gravel, fill materials).
	  ▪ Political, ethnic, or tribal boundaries and locations of government centers.
	  ▪ Structures—availability of existing structures and local facilities and infrastructure that can help sustain base camps.
	  ▪ Capabilities—ability of local economies and local businesses and laborers to support base camps.
	• Organizations within and outside of the AO that can support or affect base camps. This includes—
	  ▪ Local labor unions.
	  ▪ Criminal organizations.
	  ▪ Community watch groups.
	  ▪ Governmental and nongovernmental agencies and organizations.
	• Effects of indigenous and transient civilians on base camps (dislocated civilians).
	• Routine, cyclical, planned, or spontaneous activities that can affect base camps (holidays, elections, celebrations, demonstrations).

Legend:	
AO	area of operations
ASCOPE	area, structures, capabilities, organizations, people, and events
KOCOA	key terrain, observation and fields of fire, cover and concealment, obstacles, and avenues of approach
OAKOC	observation and fields of fire, avenues of approach, key terrain, obstacles, and cover and concealment

(FOUO content removed – figure.)

Appendix B: Spatial Components of Protests, Demonstrations, and Rallies

This appendix provides details related to section 2.3 – Spatial components.

Protests, demonstrations, and rallies have long been a vehicle for expressing political dissatisfaction. Research into the causes and outcomes of these events has primarily focused on psychological and sociological factors, as if the events occurred on a blank canvas. Over the past several decades, however, research into these types of potentially violent events has begun to include another dimension, that of space. Taking the built environment into account opens up a new avenue of research by focusing on how spatial elements of the area, meaning where these events occurred or are likely to occur, serve as attractors or detractors to protests, demonstrations, and rallies.

Sociopolitical contradictions are realized spatially. The contradiction of space thus makes the contradiction of social relations operative. In other words, spatial contradictions 'express' conflicts between sociopolitical interests and forces; it is only in space that such conflicts come effectively into play and in doing so, they become contradictions of space (Lefebvre 1991, 365).

Defining spatial environment/use of space

Typologies of space

Geographers have been defining space and place for generations. More recently, the concept of space has expanded from an indefinite area bounded in some way to a constantly shifting template within which social, temporal, economic, and political activities play out. Space can contain nodes (places) or networks (connections of places) that are shaped by and help to shape the activities within the space.

The concepts of space and place have been examined through a variety of perspectives over the years. The study of spatial systems focused on spatial arrangements of spatial structures——how human activities utilize location and how these activities spur resulting spatial interactions (Johnson 1983). Behavioral geography focuses on how an individual perceives his

spatial environment and reshapes it (Gold 1980). More useful to a discussion of effects of space and place on protests, demonstrations, and rallies is a typology developed by Lefebvre (1991). In his influential book *The Production of Space*, Lefebvre characterized space as a social product (its significance is socially produced) that serves as a tool of thought and action. The meaning of space is shaped by the predominant means of production, but it can also be a means of control or domination (Lefebvre 1991). Lefebvre categorized space into three types: the perceived space––a combination of social life and perception; conceived space––the rigorous, methodical space of cartographers, urban planners, architects, and others that work to quantify space; and lived space––a combination of both other types, but reconfigured by inclusion of individual imagination and aesthetic sense to form a signified environment. In a 1993 review of Lefebvre's *The Production of Space*, Molotch (1993, 888) provides this interpretation of Lefebvre's definition of space:

> A space is thus neither merely a medium nor a list of ingredients, but an interlinkage of geographic form, built environment, symbolic meaning, and routines of life. Ways of being and physical landscapes are of a piece, albeit one filled with tensions and competing versions of what a space should be. People fight not only over a piece of turf, but about the sort of reality that it constitutes.

In examining his concept of lived space, Lefebvre separated abstract space (commodified and bureaucratized) from concrete space (the location of everyday life and experiences). Lefebvre's typology of space has provided the major underlying conceptual perspective for many authors investigating spatial aspects of contentious politics, with most research focusing on the lived space where physical reality and symbolic meaning are integrated and influence each other. The work of Martin and Miller provides a detailed and comprehensive review of the existing literature in this field. The two authors present Lefebvre's typology through the categories of space, place, and scale, emphasizing the construction of space as a combination of social relations and structures since "space is an integral part of all social life, both affecting and affected by social action" (Martin and Miller 2003, 145).

Places are localized expressions of space that have a socially created identity that is shared to a greater or lesser extent by the inhabitants of that area. For Brantingham (2011), the persistent images we form of places

both shape activities and are shaped by them. These places contain activities such as work or entertainment, have a vernacular architecture, and unique collections of residents, shops, parks and other elements. Brantingham (2011, 201) called areas well-known to an individual as that person's "awareness space," which she defined as "places that are recognized by an individual and where an individual knows how to get to and from. In the aggregate, cities have areas that are part of the awareness space of many individuals. These areas are usually the most active within cities."

Spatial scale is also a characteristic that shapes our awareness of places. "Scale is an inextricable component of the production of perceived, conceived and lived space" as it provides dimensional boundaries for place-based activities (Martin and Miller 2003, 148). For well-known or iconic places such as Manhattan or Tiananmen Square, the areal extent of knowledge of a particular place may extend to the global scale, but the identity of a place is primarily a local social construction. Protests are always local, but they can also become regional or national as other participants are attracted by the content or representation of the contentious issue at hand. Some social issues are intertwined at several scales, such as local labor inequalities as related to national or multinational corporations. Sewell (2001) among others describe how it can be advantageous for social movement to "jump scales" from local to national in an effort to accrue more power for their cause. Jumping scales is greatly facilitated by both traditional and social media to get the word out and to publicize high-level supporters.

Temporal aspects of space

That space has a temporal dimension has been known since antiquity, as people developed processes for understanding natural cycles and applying them to spatial activities such as agriculture, hunting expeditions, navigation, and weather forecasting (Couclelis 2005). Spatial activities may vary according to the time of day, day of week, or time of year. These temporal changes both alter our spatial behavior, and reconfigure our spatial environment. The concept was developed by Hagerstrand in the early 1960s and involved the development of space-time paths that combined location and temporal data to create activity paths, primarily for individual entities (Wachowicz 2003). Hagerstrand's work was further developed by Pred (1984, 280) who presents a concept of place that derives both its form and its significance from the ceaseless changes occurring over time; "place is conceptualized partly in terms of the unbroken flow of local events." Pred

(1984) investigated the structural forces impacting individual paths by including time-allocation and scheduling precedence as a factor that imposed restraints on activity.

Through a temporal framework, the ability to protest may not depend on distance or proximity, but rather tradeoffs between time and objectives. The ability to congregate and protest depends greatly on time-distance costs. Time-distance costs is how long it takes to satisfy a goal, whether that goal may be holding protests, recruiting, or reaching or evading authorities. These goals are highly dependent on transportation and communication technologies, and the advancement of these may reduce time-distance costs in protests. However, it is seen that peripheral locations are not as effective in mobilizing, due to the physical nature of protesting in a location during a certain time in space. Stillerman (2003) described how Chilean copper strikers were able to achieve their protest aims more feasibly because their spaces of work and residence were within a few miles and situated within a commune of Santiago. In contrast, coal miners in southern Chile had significantly larger time-distance costs due to remote satellite towns and large distances to the nearest major city. Communication technologies can extend the knowledge to protest, but not necessarily the ability.

Spatial impacts on protests, demonstrations, and rallies

Sociocultural understanding of space and assignment of meaning

Space is more than a physical reality; it is also a container for socially and culturally related meanings. Space is understood through a cultural lens, and individual places may carry multiple meanings reflecting different cultural associations. Meanings can arise from traditional uses of space, such as religious complexes or college campuses. Meaning can also be created from the usurpation of traditional uses, such as a protest encampment in a public park or a sit-in at a lunch counter. Sewell (2001) stressed the malleable nature of these meanings, depending on the needs or perspectives of those utilizing the space. While spatial structures can constrain human behavior, humans are simultaneously creating, defining, and re-creating spatial structures and assigning meanings that shape behavior in that space (Sewell 2001). The shifting nature of meaning is also examined by Endres and Senda-Cook (2011) who defined place as rhetoric, with users associating preexisting meaning with a particular place, then reconstructing meaning repeatedly through behaviors performed in that place. As such, place

has both physical and metaphorical aspects, a definition in line with Lefebvre's classifications. Endres and Senda-Cook illustrated their concepts through the example of Alcatraz Island. Long associated with a prison, the abandoned facility was occupied from November 1969 to June 1971 by the American Indian Movement. Through their protest, the occupiers sought to reconstruct the meaning of the place, by trying to shift it from federal property to land belonging to the indigenous, thereby using place meaning as a "tactical act of resistance" (Endres and Senda-Cook 2011, 258 and 269). Specific place meanings can attract protests either for emphasizing the predominant meaning (saving a beloved historic building) or for an opportunity to reconstruct the place's meaning into something else, at least temporarily.

There are many examples in the literature of specific spaces or places that describe historical changes in their sociocultural meanings. Allegra et al. (2013) include the role of history in identity creation; that cities should be seen as an area of social and historical processes that create environments of tension and inequality, potentially leading to protest. Cybriwsky (2015) examines the impact on Kiev, Ukraine's historic Independence Square of the 2013-14 protests that ousted President Viktor Yanukovych after he declared a closer alignment with Russia instead of Europe. The square was created in 1876 in association with the new city administration buildings on the site. The name of the square changed several times over the years to reflect cultural meanings for the occupiers as Kiev was occupied by the Soviet Union and Germany, and the square served as a place for government celebrations. After Ukrainian independence in 1991, the square was aptly renamed Independence Square and became associated with nationalism, protection of the homeland, and emergence from oppressions of the past. As such, it was the central site of protests against government policy perceived as threatening to national solidarity. The square became so associated with these protests, that the Ukrainian word for public square (maidan) was utilized as a call to protest as in "come to the square" (Cybriwsky 2015, 270). After several months of occupation of Independence Square by protesters, government forces intervened violently to disperse the crowd, resulting in over 100 fatalities and the President fled. The subsequent funeral services were held on Independence Square, thus again shifting the square's meaning from one of protest to one of memorialization and remembrance. Other studies that have focused on this type of meaning transformation of a space include Salmenkari (2009: the Plaza de Mayo in Buenos Aires from government rallies to a site of resistance),

Sewell (2001: Tiananmen Square from government rallies to expression of democracy to a site of martyrdom), and Ismail (2013: the transformation of residential quarters in Damascus to reflect government priorities and political parties).

That places can possess identities of inequality has been addressed multiple times in the literature (McCann 1999, Stangl 2010, Martin and Miller 2003, among others). "People can see inequality inscribed in the landscapes of their daily lives" (Martin and Miller 2003, 146). Allegra et al. (2013) discussed how historical and social processes can create environments that are seen in terms of tension and inequality, and that urban protests related to perceptions of inequality play a role in initiating change. The perception of inequality as an inherent characteristic of particular places increases the likelihood of those places becoming sites of protest. If the level of perceived inequality increases at a rapid rate or passes a certain level, the site becomes increasingly symbolic of that inequality. A good example of this is provided by Schmidt and Babits (2014, 79) who described "contested public representations of occupation" in the Occupy Wall Street movement of 2011. Arising from several years of deep recession, the protest over economic inequality as controlled by "the 1%" of wealthiest Americans was held in a site near the most representative symbol of this inequality, the financial institutions of Wall Street.

The built environment

A majority of the research on spatial aspects of protests, demonstrations, and rallies is focused on cities. This seems almost definitive of these types of gatherings, as a critical mass of people are needed for the presentation of alternative social and political ideas expressed in this manner. In addition to the symbolic nature of protest sites, their physical realities impart advantages and disadvantages, access and barriers, likely and unlikely options. It is instructive to investigate the physical characteristics of cities as a basis for understanding how people interact in their lived space with regard to the location of protests.

One of the ways in which a city can be broken into components for analysis is presented by Nejad (2013) in a discussion of the ways that physical urban spaces impact crowd behavior. According to Nejad, analyses of protests often focus on how the urban environment is signified, not how the city is laid out relationally. The space syntax topological technique is used to analyze the spatial structure of a city, and the technique has as its basis

the idea that buildings and cities are ordered together as a whole and those relationships can reveal how cities function (Hillier and Hanson 1984). Therefore, the complexity of cities can be analyzed through its interdependent parts; specifically, the parts themselves and the relations between parts. The Nejad article provides a methodology based on the two measures of connectivity (connections between nodes) and depth (number of steps between two nodes). As a variant of central place, this methodology works to define centrality and to quantify interconnection and access. Using Tehran as an example, Nejad (2013) examined the role of urban street networks and the integration of highly accessed sites in the development of crowds in the central commercial areas. The area with the most crowd development (the most integrated area for urban movement) was along a major commercial city street with public squares at each end and along the street's axis that also held symbolic meaning.

On a different scale, the built environment can be analyzed as a series of zones or neighborhoods. These areas of cities are often characterized by high residential density. Zhao (1998, 1497) described the effect of density on population as having an effect on social behavior, as "other factors being equal, the closer a number of people live together (in both physical and functional terms), the greater the chance of unintentional contacts and active group making." Rookey, Christian, and Van Dyke (2005) highlighted the role of the built environment in an investigation of student protests on campuses in the United States, bringing together an analysis of specific protest events, location, and collective political action. Following on to Sewell (2001) and Zhao (1998), the authors stressed the importance of the built environment, stating that it creates and shapes social interaction, provides the possibility of protest, and impacts spatial routines. These spatial routines, undertaken nearly universally, assist in the formation of social networks such as familiarity with regulars at a local coffee shop that forms part of a weekday commute. When the spatial routines of large numbers of people coincide, the place of coincidence can be the site of protests. In particular, campuses that had gathering places such as a quadrangle and had sufficient population density for ideas and information to spread, have historically experienced more protests.

The land use demarcations of a city are part of the built environment and can help shape social and spatial behavior, particularly when the uses have physical manifestations that serve as distinguishing characteristics. The

most relevant land uses for this discussion are public space and private space.

Public space

In democratic societies, public space provides an opportunity space for protests and demonstrations. Schmidt and Babits (2014, 80) utilized this civic framework to discuss collective occupation of public sites in the United States for political, social, and economic dissent. The sites are seen as belonging to "the people," where they can congregate to "protect their common interest and produce an outlet for dissent against the government". Often, public space as designed by planners (e.g., Lefebvre's conceived space) conflicts with the lived experience of the public and protestors. As a result, "violent clashes arose when protestors controlled space in a way that benefited their political cause but deviated from the ways in which that space had been used in the past" (Schmidt and Babits 2014, 82). Public space is not solely the representation of government as are courthouses, police stations, and congressional buildings. Public space is also the streets, parks, playgrounds, sidewalks, and parking lots that shape experiences as people move through and utilize city sites in their spatial routines.

Private space

Schmidt and Babits (2014, 80) also look at the concept of private space, characterizing it as spaces "of production and consumption owned by individuals or corporations." That ownership is key; it is what enables opponents of protest to physically constrain sites of protests. The authors noted that conversely, private sites (particularly corporate headquarters or sites of production) can be attractive to protestors decrying perceived corporate injustice. Private space is often contested space as the socio-economic characteristic of the owners may not reflect those of the users or of the local inhabitants. In a discussion of racialized geographies, McCann (1999, 164) described the private spaces of downtown business districts as "exclusionary territories dominated by White, middle-class males." Private space in the United States has not typically been used as a gathering site for protests, although demonstrations and protests sometimes move through them.

Increasing privatization of public land

The line between public and private space is becoming increasingly blurred. McCarthy and McPhail (2006) provided a detailed discussion of the increasing privatization of public space, contending that public fora are shrinking in number, are more difficult for the public to access, and are no longer popular for gatherings. Places move from public to privatized (or at least no longer allowing public access fora), because regulations proliferate that govern "acceptable" activities in these types of places. Those regulations say that protests must be permitted, only specific areas can be utilized, and plans must be submitted in advance, all of which serve mostly to inhibit the use of public spaces as sites of protest or dissent. Access is also restricted through the takeover or management of public space by private interests, such as where "public sidewalks are privatized in gated communities, and also, to some extent, in downtown Business Improvement Districts" as well as public parks being operated by private concerns (McCarthy and McPhail 2006, 229). Some of these formerly public areas may still be accessible, but behavior is often controlled by private security personnel. At the other end of the spectrum, there are many instances where space is used in a public manner when in fact, it is privately held. Shopping malls, sports arenas, and concert halls are the preferred locations for large gatherings of people, not public plazas or civic structures. These areas do not serve as public places for protests, however, as they are private facilities and are not required to allow the exercise of free speech.

While privatization of public space may be occurring in the United States, the process may not be in place elsewhere. Salmenkari (2009) highlighted two examples from other parts of the world, and he did not find a lessening of public protest locations or of increasing restriction on semi-public areas. The example of Seoul, Korea, described the city as having been built on the traditional Chinese model, with no public plazas and a dense web of narrow streets. The area around the presidential palace was closed off, but the main roads served as both vehicular and pedestrian thoroughfares, with government and commercial uses. The popular culture is one of consumerism, and the commercial spaces are the gathering places and traditional sites of protest in addition to government sites. According to Salmenkari (2009, 249), protests at commercial sites are attractive because "events did not take place in a politically contested zone, [so] authorities had little interest in them," and the property owners do not discourage them. Buenos Aires, Argentina, was constructed on the European model, with wide boulevards, square and other public open spaces,

and monumental public buildings visually associated with government and politics. These areas have symbolic significance as well, and combined with an active street life that utilizes public areas, they provide the traditional and continuing venue for protests. Demonstrators want to directly confront the authority in charge of the issue being protested.

How people interact with the built environment

The built environment shapes and is shaped by the desire to protest and the opposing desire to prevent protest. This is most likely to occur in urban areas due to higher public visibility to local and broader audiences; density of population; ease of communication, access to sites, concentrated location of government facilities, headquarters of businesses and unions, etc. There are many sociocultural drivers associated with place, space, and the built environment that impact where protests occur and how the space around the protest site is utilized. Three of these drivers are most often discussed in the literature: inequality, power, and areas of population contention. An article by Allegra et al. (2013, 1679) draws on previous work by multiple authors to combine these areas of focus as a useful perspective on protest, stating that:

> In the first place, from an urban social movement perspective, the city is mainly seen as the environment that creates the structural conditions for dissent to emerge and be expressed. There is in fact a long and established tradition of enquiry which sees the city as a place of alienation marked by poverty, segregation, lack of security, violence, repression and the loss of communitarian ties, with these structural features automatically producing the potential for social struggle.

Inequality

Inequality in many forms is expressed in the built environment, such as the following: crowded slums versus spacious housing, narrow alleys versus wide boulevards, concrete playgrounds versus grassy parks, and a multitude of gates, signage, checkpoints, and other barriers that serve to separate rich and poor. Various types of physical barriers are also utilized to enforce separation based on race, social and educational status, and other aspects of social differentiation. Martin and Miller (2003, 146), describe the relationship of the disenfranchised as "inequality inscribed in the landscapes of their daily lives."

In an investigation of protests' urban geography, Salmenkari (2009) noted that center-city workers in Buenos Aires often lived in the poorer barrios on the outskirts, yet would travel back to the affluent city center for protests. In Jakarta, Indonesia, the poor protest at the most luxurious spaces in the city, including the Presidential Palace and an upscale hotel. Padawangi (2010) describes the use of these areas for protest as the poor redefining the exclusivity of these spaces, representing a broad class struggle in the city. In the case of South Africa, a history of racial separation was made physical in the creation of "homelands" and suburban townships for the Black population as a means of exclusion. According to Jelly-Schapiro (2014), resistance arose in the urban townships as the local population co-opted their townships as places of autonomy with their own society and defended that society against intrusion. Jelly-Schapiro quotes from Bozzoli (2004, 69) as follows: "Confronted with borders designed to separate and confine – to keep Black people in, the rebels transformed the township's boundaries into metaphorical and at times actual barricades designed to keep outsiders out."

Power

Much of the examined literature on inequality as related to protests, demonstrations, and rallies include this driver of resistance as one of several manifestations of a larger struggle to possess and apply power over a population or its resources, including the built environment. Sewell (2001) defines power as control over people and territory, with carefully marked and monitored boundaries; some form of policing is required to exert and maintain this control. McCarthy and McPhail (2006) discuss the role of police in determining the locations that protestors are allowed to gather. The choice by police of where to place barriers, the demarcation of sanctioned protest areas, and the control of transit routes all serve to demonstrate the established law enforcement's power against the protestors. Salmenkari (2009) describes displacement tactics utilized by police in Buenos Aires, who enforced no protest zones with riot fences and established permanent no-protest zones around the Congress and the Presidential Palace. In Seoul, police riot lines and police buses are used to create mobile boundaries. It is possible, however, to remove control from the policing agents through protest. During the revolution in Cairo, the backstreets of the crowded old quarters of the city contained police stations, which were symbolic of oppression to the inhabitants (Ismail 2013). Many were burned by the protesters, enabling the protests to continue by disarming the police.

Another aspect of power inequality is expressed in the choice of protest location. McCarthy and McPhail (2006) create a duality where protests that target private actors occur in private spaces while protests that target the state occur in the limited public forum space. Zhao (1998) describes students at Beijing University as initiating their prodemocracy protests on university grounds but then moving to public streets. During the protests over the 2009 presidential election in Tehran, increasing government restrictions on protests resulted in fewer central areas that the protesters could access (Nejad 2013). The protesters were forced to peripheral residential areas, effectively lessening their overall impact.

Areas of population contention

Social spaces reflect the societies that create and utilize them. When the societies sharing the same or adjacent built environment spaces are in some type of conflict (e.g., politically, ethnically, philosophically, militarily), the opportunity for protest is increased (Martin and Miller 2003). According to Horowitz (2001), the strongest riots often take place where there is the most support for the marginalized group or where competition between different groups is the strongest. Ismail (2013) presents the cases of Cairo and Damascus in terms of shifting politics and manipulation of contentious populations. In Cairo, the Arab Spring protests were a conflict between the urban populace and the state as represented by the police. The old quarters of the city were becoming increasingly disassociated from the government through squatter activity and reduction of social services. The resulting sense of autonomy among the residents was a challenge to the state, and the challenge was met with attempts at retaining government control thought police activities. Confrontations often occurred in the old quarters—the areas of oppression and resistance. In Damascus, neighboring areas in the city were manipulated by the government to fragment and diffuse potential dissent. Military families were settled in a restive quarter of the city to create a buffer and to fragment opposition. The Syrian regime also attempted to disperse dissent in another quarter of Damascus by spatially changing the area through construction of wide roads and high-rise buildings.

Mobility/transport factors in the built environment

The urban built environment can vary from a planned arrangement of wide roads, plazas, parks, and controlled building to a dense, narrow, haphazard organic development. As the success of protest movements often

rests on the size of the crowd, it is essential that protesters have access to the protest site. In some cases, the protest site is inherent to the protest population. Zhao (1998) describes how student protests in Beijing began at the campus of Beijing University on the day before the crowd moved to Tiananmen Square. More common, however, is for protests to either be set for a particular location in a central part of the city, or to be a march culminating in a central place. Both of these spatial types of protests require movement in space.

According to Rookey, Christian, and Van Dyke (2005), movement has a time-distance cost, the amount of time it takes to get from one place to another. This cost can influence the choice of protest location, as larger crowds can usually be assembled if the time-distance cost is lower. The time-distance factor also affects the level of attention or coverage afforded to a protest by the media, with urban areas again offering lower costs due to proximity. Sewell (2001) also includes time distance in a discussion of access to demonstration locations, focusing on the "everyday mobility" that brings large numbers of people together, such as weekly markets as a site for food riots, or worker demonstrations that occur near bars where the workers routinely go after work. In these cases, the time-distance cost is minimal. The protest camp is another type of protest with an initial time-distance cost that is mediated over time. Once at the site, demonstrations occur over an extended period of time when little additional movement is required, as people congregate in areas of convergence (Frenzel, Feigenbaum, and McCurdy 2014).

Due to their very nature as means of transit, many streets, walkways, and railways offer paths of mobility to protest sites, particularly as they tend to converge in central urban areas. Nejad (2013) discusses the most commonly used street networks in Tehran, and speculates on their utility in predicting their likelihood of drawing pedestrian protests. While major transport arteries can be useful in gaining access to a protest site or participating in organized marches, they have disadvantages as well. Sewell (2001) states that areas with high building density and narrow streets provide protection for demonstrators as they can easily be blocked by the crowd, whereas wide boulevard-like streets provide less cover and allow access to those suppressing the demonstrations. According to Sewell, the intimate knowledge of an area's spatial structures can be an advantage to the people who created them and can use them as a means of resistance.

Stillerman (2003) investigated the mobility issues involved in a 1960 Chilean metalworkers strike. The protests occurred in the San Miguel area of Santiago, where the factory subject to the strike was located. The protesters utilized the built environment in their actions against the factory and police. Most of the strikers lived in concentrated housing blocks relatively near the factory, providing them with low time-distance costs. This housing enabled the strikers to retaliate against the strikebreakers, while using local refuges for protection from police. The factory was only a few miles from the political center of Santiago, so the strikers also had easy access to a populated area for conducting marches.

In comparing urban morphology as it related to access to protest sites in San Francisco and Los Angeles, Stangl (2010) determined that San Francisco's small city block size, higher density of buildings and population, effective regional rail network, mixed urban uses, and pedestrian-friendly transit routes provided much greater access for marches and demonstrations in public areas than was the case in Los Angeles.

Conclusions

The following lists of built environment elements that either attract or detract the possibility of protests, rallies, and demonstrations is gathered from the literature discussed above. It is not a definitive list, but meant to enable efficient examination of the specific roles of these elements, and other elements that may also be involved in these types of political, economic and social events. Of particular note is the fact that many of the attractors can be altered by the authorities to become detractors intended to prevent or discourage protests, rallies, and demonstrations.

Attractors in the built environment

Spatial attractors

- Large central commercial sites
- Dense, multistory apartments
- High levels of marginalized populations concentrated in particular areas
- Spatial patterns and routines that are not conducive for community policing
- Large number of people in a particular place
- Large open spaces at intersections of main transit ways

- Public squares or plazas
- High-level government buildings (palaces, parliaments, police/military headquarters, political party headquarters, embassies, etc.)
- High-level private buildings (corporate headquarters, banks, stock exchanges, elite residential areas, etc.)
- Historical or religious sites or centers
- Familiarity with the protest space
- Familiarity with transit routes and ease of access
- Linkage between features/protest routes
- Sidewalks or walkways that are open and accessible to pedestrians
- Open public land such as parks, playgrounds, and parking lots
- Places that provide physical access to directly confront the symbols of authority

Temporal attractors

- Low time-distance costs
- Fits mass transit schedules
- Times when the group is already present near the protest space
- Protests that occur at regular intervals or schedules

Detractors in the built environment

Spatial detractors

- Low-density residential or individual units
- Improvised barricades or borders
- Small- and medium-sized streets defendable against protests
- Subdivided public areas – fenced off, barricaded, policed
- Wide central boulevards as "no man's land" – hard to cross, easy to police
- Large public squares and other spaces can be "filled" with street furniture (benches, bollards, fountains, planters, etc.) that inhibit large crowds
- Space too constrained – either by physical borders or by barriers erected on the site
- Space without strong symbolic elements of authority
- No linkages between protest spaces
- Streets with police roadblocks to turn back protesters
- Formerly public space that has been privatized and controlled (residential areas, parks, walkways)

Temporal detractors

- At inconvenient times for travel to protest site
- At times when possible participants are not in the area
- Infrequent mass-transit schedules
- High time-distance costs

Literature cited in Appendix B

Allegra, M., I. Bono, J. Rokem, A. Casaglia, R. Marzorati, and H. Yacobi. 2013. "Rethinking Cities in Contentious Times: The Mobilisation of Urban Dissent in the 'Arab Spring'" *Urban Studies* 52(11): 1675–1688. Doi:10.1177/0042098015590050.

Bozzoli, Belinda. 2004. *Theatres of Struggle and the End of Apartheid*. Athens: Ohio University Press.

Brantingham, Patricia. 2011. "Crime and Place: Rapidly Evolving Research Methods in the 21st Century." *Cityscape: A Journal of Policy Development and Research*, special edition on Crime and Urban Form, 13(3): 199–203.

Couclelis, H. 2005. "Space, Time, Geography." In *Geographic Information Systems: Principles, Techniques, Management and Applications*, 2nd edition. Hoboken, NJ: Wiley.

Cybriwsky, Roman. 2015. "Kyiv's Maidan: From Duma Square to Sacred Space." *Eurasian Geography and Economics* 55(3): 270–85. Doi:10.1080/15387216.2014.991341.

Endres, Danielle, and Samantha Senda-Cook. 2011. "Location Matters: The Rhetoric of Place in Protest." *Quarterly Journal of Speech* 97(3): 257–82. Doi:10.1080/00335630.2011.585167.

Frenzel, Fabian, Anna Feigenbaum, and Patrick McCurdy. 2013. "Protest Camps: An Emerging Field of Social Movement Research." *The Sociological Review* 62(3): 457–474. Doi:10.1111/1467-954X.12111.

Gold, J.R. 1980. *An Introduction to Behavioral Geography*. Oxford: Oxford University Press.

Hillier, Bill, and Julienne Hanson. 1984. *The Social Logic of Space*. Cambridge, England: Cambridge University Press.

Horowitz, Donald. 2001. "Location, Diffusion, and Recurrence." In *The Deadly Ethnic Riot*, 374-422. Berkeley and Los Angeles, CA: University of California Press.

Ismail, Salwa. 2013. "Urban Subalterns in the Arab Revolutions: Cairo and Damascus in Comparative Perspective." *Comparative Studies in Society and History* 55(4): 865–894. Doi:10.1017/S0010417513000443.

Jelly-Schapiro, Eli. 2014. "Occupation against Occupation: Space and Anticolonial Resistance." *Transforming Anthropology* 22(1): 46–52. Doi:10.1111/traa.12024.

Lefebvre, H. 1991. *The Production of Space*. Cambridge, MA: Blackwell, 1991. (First published as La Production de L'espace, 1974).

McCarthy, John, and Clark McPhail. 2006. "Places of Protest: The Public Forum in Principle and Practice." *Mobilization* 2(2): 229–247.

Martin, Deborah, and Byron Miller. 2003. "Space and Contentious Politics." *Mobilization* 8(3): 143–56.

McCann, Eugene J. 1999. "Race, Protest, and Public Space: Contextualizing Lefebvre in the U.S. City." *Antipode* 31(2): 163–184.

Molotch, Harvey. 1993. "The Space of Lefebvre." *Theory and Society* 22(6):887–895.

Nejad, Reza Masoudi. "The Spatial Logic of the Crowd: The Effectiveness of Protest in Public Space." *International Journal of Islamic Architecture* 2(1): 157–78. Doi:10.1386/ijia.2.1.157_1.

Padawangi, Rita. 2010. "From Backstage to Frontstage: Place-Making, Protests and the Empowerment of the Urban Poor." Conference Papers – American Sociological Association 2048, SocINDEX with Full Text, EBSCOhost (accessed October 28, 2015).

Pred, Allan. 1984. "Place as Historically Contingent Process: Structuration and the Time-Geography of Becoming Places." *Annals of the Association of American Geographers* 74(2):279–297.

Rookey, Bryan, Leah Christian, and Nella Van Dyke. 2005. "The Influence of Space on Student Protest." In Conference Papers of the American Sociological Association. 2005 Annual Meeting in Philadelphia, 1–20.

Salmenkari, Taru. 2009. "Geography of Protest: Places of Demonstration in Buenos Aires and Seoul." *Urban Geography* 30(3): 239–60. Doi:10.2747/0272-3638.30.3.239.

Schmidt, Sandra J., and Chris Babits. 2014. "Occupy Wall Street as a Curriculum of Space." *The Journal of Social Studies Research* 38(2): 79–89.

Sewell, William H. Jr. 2001. "Space In Contentious Politics." In *Silence and Voice in the Study of Contentious Politics*, 51-88. Cambridge, England: Cambridge University Press.

Stillerman, Joel. 2003. "Space, Strategies, and Alliances in Mobilization: The 1960 Metalworkers and Coal Miner's Strikes in Chile." *Mobilization* 8(1): 65–85.

Wachowicz, Monica. 2003. *Object-Oriented Design for Temporal GIS*. Boca Raton, FL: CRC Press.

Zhao, Dingxin. 1998. "Ecologies of Social Movements: Student Mobilization during the 1989 Prodemocracy Movement in Beijing." *American Journal of Sociology* 103(6): 1493–529.

Additional literature reviewed for Appendix B

Ford, Matt. 2004. "A Dictator's Guide to Urban Design." *The Atlantic*.

Jansen, Stef. 2001. "The Streets of Beograd. Urban Space and Protest Identities in Serbia." *Political Geography* Vol. 20(1): 35–55.

Johnston, R.J. 1983. *Geography and Geographers: Anglo-American Geography since 1945*, 2nd Edition. London: Edward Arnold.

Kelley, Strawn. 2008. "Validity and Media-Derived Protest Event Data: Examining Relative Coverage Tendencies in Mexican News Media." Mobilization 13(2): 147–164.

Khatiwada, Lila Kumar. 2014. "A Spatial Approach in Locating and Explaining Conflict Hot Spots in Nepal." *Eurasian Geography and Economics* 55(2): 201-217. Doi:10.1080/15387216.2014.956135.

Makhoul, John; Kubala, Francis; Schwartz, Richard; and Weischedel, Ralph. February 1999. "Performance Measures for Information Extraction." In *Proceedings of DARPA Broadcast News Workshop*, held in Herndon, VA.

Marom, Nathan. 2013. "Activising Space: The Spatial Politics of the 2011 Protest Movement in Israel." *Urban Studies* 55(3): 2826–2841. Doi:10.1080/15387216.2014.991341.

Stillerman, Joel. 2006. "The Politics of Space and Culture in Santiago, Chile's Street Markets." *Qualitative Sociology* 29(4): 507-30. Doi:10.1007/s11133-006-9041-x.

Stark, Margaret J. Abudu, Walter J. Raine, Stephen L. Burbeck, and Keith K. Davison. 1974. "Some Empirical Patterns in a Riot Process." *American Sociological Review* 39(6): 865–876.

Zhao, Dingxin. 2003. "Organization and Place in the Anti-US Demonstrations after the 1999 Belgrade Embassy Bombing." Conference Papers – American Sociological Association 2003 Annual Meeting held in Atlanta, GA, 1–39. Doi:asa_proceeding_8560.PDF.

Appendix C: Event Models

This appendix describes additional details about event modeling and analytical methods related to section 4.3.

In geosocial analysis, there are many ways to define an event, but no unified definition has been agreed upon (Subsection 2.1 provides a possible one). While events are often thought of as high-visibility developments, as in a political protest, they can be much finer-grained. The detection of new bacteria, often without any health consequences, is an example. What events do have in common, however, is the need to be well-described and correctly located in a physical region. Events can be described by the following components:

- *Entities* represent the actors involved in the action, such as protesters and police officers. Commonly, entities denote people and organizations and less frequently, they refer to tangible objects as well as intangible concepts or ideas.

- *Features* describe intrinsic characteristics of an entity. In most cases, features remain static, although changes can arise over extended periods. Features could describe a suspicious backpack left behind or the description of a looter. Features not only enhance the analysis, but also serve to differentiate entities and provide behavioral clues.

- *Interactions* are the relationships shared among entities. In terms of violent events, interactions often have a physical connotation such as when looters break windows. Interactions, however, can represent any action that affects one or more entities. Under different domains of analysis (e.g., financial, social, medical, political), they can vary widely.

- *Influence* extends interactions by means of cause and effect. While an interaction describes an observed action, influence measures the extended repercussion due to that action. The murder of a journalist (i.e., the interaction), for example, may have a chilling effect on people reporting corruption (i.e., the influence). Influence is not only challenging to understand, but it can also be misleading since

it is not always clear if an observed effect was truly generated by the suspected cause.

- *Time* provides a sequential view of entities, their interactions, and their influence. As events evolve, entities go in and out of sight while their interactions define new contexts. Time, then, becomes essential as a means to compartmentalize the different contexts that can become more intelligible when viewed as separate units. A political protest, for instance, may be modeled as a collection of interactions (e.g., looting and police intervention) broken down into several snapshots, which may make the analysis more concise.

- *Locations* can be defined for both entities and interactions. From a semantic perspective, however, locations are better suited when operated on the ongoing interactions. The reason is that entities can have many locations (even in the same time window), while interactions tend to be more atomic concepts. In the case of looting, for instance, its location can be viewed as a few city blocks (one polygon), or as the changing locations of entities as they enter, move around, and exit the scenario. Locations are often challenging to interpret, and they are discussed further below. Figure C1 provides a visual representation of the model discussed above. It shows the hypothetical interaction between two entities observed in a geographical location and represented along timestamps [t], [t+1], and [t+2]. Each entity is described by a notional feature set F and as time elapses, the sphere of influence generated by the interaction grows in space.

Figure C1. Event model: entities, features, interactions, influence, time, location (ERDC-CERL).
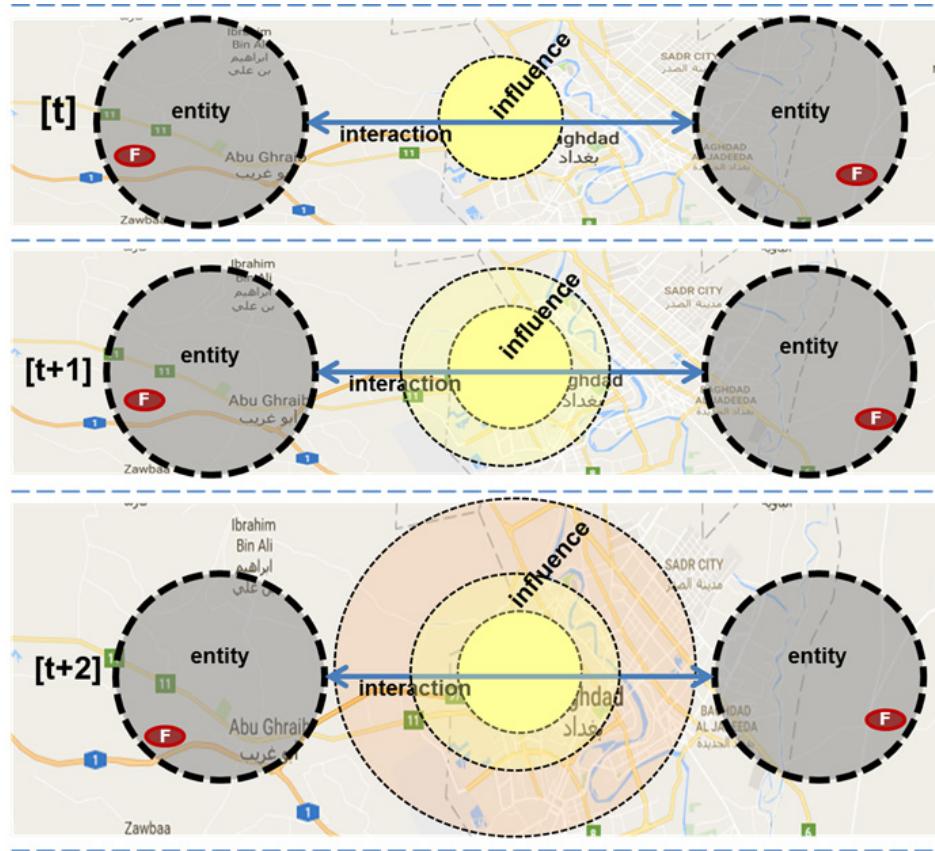
Representing locations

The aspect of location raises several questions as to the most appropriate approach to defining an event's place——whether it is a single point, a line, or an extended region (i.e., a polygon). Once a location has been established, one should question the applicability of data representation in a raster or vector format, both of which present advantages and challenges. For the two tasks just mentioned (i.e., determining the location and representing it in an appropriate format), the notion of uncertainty is guaranteed to manifest itself in various ways, and that uncertainty must be accounted for in order for results to become explainable.

Determining location of an event

A seemingly simple question, but with major computational implications is how to determine the most appropriate location for an event. While one may be tempted to pinpoint the most obvious area, this choice can quickly become unclear. Take, for instance, a speech in front of the presidential

palace, attended by thousands of people stretching down several blocks. If the event is defined as the "speech," then its location can simply be a point representing the latitude and longitude of the presidential palace. Alternatively, one may be more interested in the attending crowd than in the speech itself. In this case, the event location should encompass the crowd and be represented by a polygon of the street blocks. It then becomes more apparent that for events, location is not a static feature, but rather a function of the context in which the event must be understood. The choice of representation must be considered in light of the following factors:

1. A *point* is the most basic representation within a grid or map. It is attractive in terms of its low storage requirements, and it demands fewer computing cycles than other formats. On the other hand, its interpretation can often be misleading when the application requires accuracy. The "speech" example above, for instance, would look strange if location were to be represented by a point behind the presidential palace, and reported in the news as such.
2. *Lines*, viewed as a collection of points, are applicable when the context of analysis requires sequential continuity. Riot police, for instance, could be represented as a line stretching along the protesting crowd. It clearly requires more storage than just points, but computational complexity should still be manageable. Lines become challenging to work with when the underlying data does not provide a clear sequence (i.e., when the data is incomplete and continuity must be estimated or assumed.
3. *Polygons* represent a collection of lines and provide the most opportunity for accuracy, but they come with high processing costs. Polygons could enclose both the "presidential palace" and the "street blocks" holding the crowd. A significant problem with polygons is that boundaries are not always clear due to missing data. Scaling them up risks the addition of empty space, while scaling them down might miss important points. Polygon data require more storage space, more computational power for processing, and effective methods of spatial indexing.

To add complexity to the notion of location, the most appropriate representation may not even be the one that makes the most sense, but the one that is most feasible for the use in question. Data uncertainty, which can also include ambiguity, noise, and errors, can impact results significantly and is discussed below.

Aspect of uncertainty

Arguably, points are the most flexible location format in that they can also represent lines and polygons. They can be used whenever the application is more concerned about the existence of an event (e.g., that a protest indeed took place) than with the details of that event (e.g., how far did the protest stretch?). It should be noted, however, that representing lines and polygons as points often means loss of information, as a single point would not be able to encode all necessary locations such as the riot police and the entire crowd. This is a type of uncertainty introduced by design for the sake of efficiency and can be managed with relative ease.

A more pressing concern relates to incomplete, missing, or incorrect data. In many cases, this concern indicates flaws in the data collection process due to human error, sensor failure, data corruption, or any combination thereof. At other times, uncertainty is intentionally caused, such as when data is "stripped clean" due to privacy concerns. Regardless, this type of uncertainty must be documented as uncertainties may play a role in understanding the final results of an analysis. Incomplete or missing data often requires the analyst to fill in the gaps on a manual basis, a very costly process. Current research has proposed several techniques to automatically estimate missing values in a dataset, as listed below:

1. *Expectation-Maximization* (EM) tries to find the maximal likelihood of a parameter in order to estimate a missing value (Dempster, Laird and Rubin 1977), a task that would be possible if the data were complete to begin with. Suppose that the probability that a protest will cover "n" city blocks depends on the number of participants, which you are trying to estimate. Given that data may only be available for previous unrelated protests, this would be a difficult task. EM looks into the known data and iteratively makes statistical guesses, using a wide array of possibilities. Thus, EM would calculate the probability of 100,000 people in 3 city blocks, 200,000 in 4 city blocks, etc. The highest probability would be the estimate for the missing value.
2. *Single Value Imputation* allows the system to fill in the missing information with plausible values (Kim and Curry 1977). The analysis would then continue as if the data were originally complete. One problem with this approach is that bias may be introduced, which changes the distribution of the data and could yield misleading results. Nevertheless, it maintains all cases in place, allowing the analyst to make a judgement as to the validity of the results.

3. *Interpolation* provides a method to estimate a missing value when given a known evolution of facts. If, for instance, a political protest grows by "x" number of participants every year, then it would be safe to assume that in the next year, it will grow by "x" participants again. This assumption can obviously be flawed, but it is often a best-faith estimate. More sophisticated approaches employ machine learning techniques and supervised methods. When the increase is constant, the interpolation is linear. By applying a weighted average, one can slow down the growth rate of the estimate if there a belief that further increases cannot attain realistic results. Alternatively, splines can be introduced so that the estimate will always go through a control point. This would be the case when the analyst is positive that the crowd will hit at least 500,000 participants at year five, though he/she cannot guarantee this would be a possible bound.

Uncertainty also affects the analysis of an event in terms of its locations. In some cases, a location may not be available. In others, several locations are mentioned, which may create ambiguity. Consider a statement which describes "a protest against the President, who was touring Japan and the Philippines." While the true location of the protest is not mentioned, a geoparsing tool may incorrectly place it in Japan or the Philippines. Social media posts are notoriously prone to such situations. Tweets, for example, very frequently do not mention any location. In limited cases, they may be GPS-enabled with the latitude and longitude of the issuing smartphone. In others, only the location of the user's account is made available. None of them is guaranteed to refer to the true location of the event being described. Often, automated systems incorporate a set of heuristics in an attempt to pinpoint the correct location. A simple approach is to select the first available location as the legitimate one. This approach is based on the notion that people often speak about one fact (or maybe very few) at a time, and picking the first location would have good odds of accuracy. More sophisticated techniques look for hints, such as well-known places, that can lead to the location in question. Therefore, if the geoparsing tool identifies "Buckingham Palace" and "The House of Commons," chances are the protest took place in London.

Location accuracy becomes relevant when the analysis requires the computation of distance between entities. When the two entities are encoded as points, the computation is either a straight line or a road path. In the

previous example, a point calculation would suffice to establish the distance from the "Police Chief" to the "President" when the "riot broke out." Computing distances from points to lines, points to polygons, or lines to polygons, however, requires extra considerations, as seen in Figure C2 and explained by the points below:

- The distance from a point (the Presidential Palace) to a line (the riot police) can be the shortest distance from the Presidential Palace to the closest police officer, as denoted by d1. Alternatively, it could be to the midpoint of the line, as shown by d2. As a more complex computation, d1 requires the system to be aware of all the distances from "PP" to every "P" in order to select the shortest one. On the other hand, d2 is simply a lookup to the middle of the line, which may be less accurate but also less computationally costly.

- The distance from a line to a polygon can also take on the closest-to-closest point approach. A common technique, however, is to select the closest point on the line to the centroid of the polygon, as indicated by d3. While applicable to regular shapes, centroids tend to appear toward the center of the polygon. In the case of irregular shapes, as in the "crowd," the centroid can simply be an arbitrary, centralized point.

Figure C2. Distance calculation.
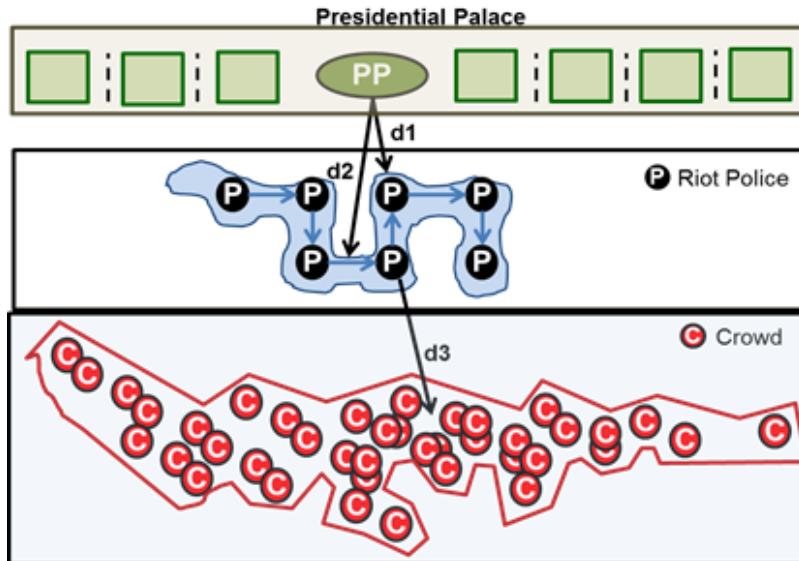
Data processing and visualization

Maps provide a friendly way of displaying the environment in a manner easily understood by the human mind. For machines, however, a map is not a single concept, but rather a collection of items described in terms of their features, locations, timestamps, and any other information that may be pertinent to the analysis. Figure C3 (A) shows a hypothetical map of the protest example mentioned previously. However unsophisticated, the map provides a bird's-eye view of each component (the Presidential Palace, the Riot Police, and the crowd) relative to one another. For machine processing, the map must be broken down into separate items to allow fine-grained control and optimize storage. In doing so, maps can be quickly retrieved on demand, and just as efficiently saved when changes are introduced. There are currently two popular encoding techniques for data representation suitable to points, lines, and polygons:

1. In a raster representation, the area of study is mapped to a grid where each cell contains information about the items on the map. An equivalent data structure would be a [n x n] array indexed on a x,y coordinate (Shirabe 2005). The content of each cell is arbitrary, and left as a design decision. Figure C3 (B) illustrates the raster visualization of (A) in an 8x8 grid by encoding each cell with one of three possibilities: a cell contains part of the crowd, the riot police, or the Presidential Palace. A nice feature of raster representation is that physical location can be implied from the feature's position on the grid. Coordinates do not have to be stored necessarily. This type of representation makes quantitative analysis quick to perform. Raster representation, on the other hand, is designed for one feature per cell, which could make the inclusion of associated data challenging. Another problem is how to select cell size which implies resolution of the data. Since each cell has a fixed size, it may become wasteful, i.e., it allocates more space than actually needed by the data. This type of encoding has a well-known scaling problem: when cells are increased, the image may become jagged since the cell does not represent the true shape of the object.
2. In a vector representation, each object is modeled as a point, line, or polygon. Figure C3 (C) displays the Presidential Palace as a point, the riot police as a line, and the crowd as a polygon. Unlike a raster data structure, the vector format has no cells. Instead, each element is positioned by its coordinates, which provides a high level of accuracy, but requires extensive storage and demands higher processing capabilities

than raster. Scaling is efficient since there is no edge distortion caused by increases in cell size.

Figure C3. Raster and vector representations.
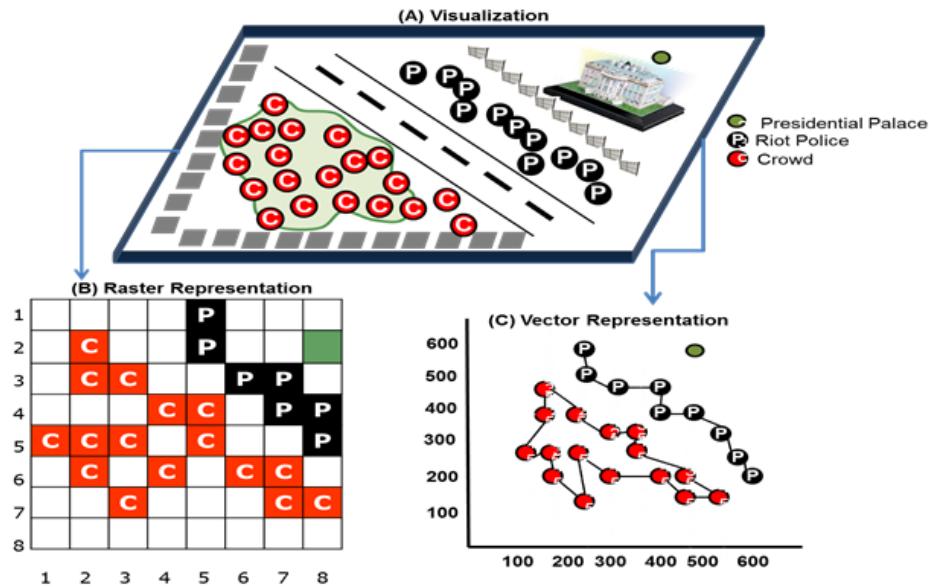
From the above discussion, it can be seen that the utilization of raster versus vector representation depends on various factors. Raster sets have been widely used for common phenomena that trend continuously (Esri 2001), in which case a cell-based approach would make sense. Spatial modeling often benefits from such setups. Vectors, on the other hand, provide precise locations that are needed in many domains. Application requirements and availability of resources often dictate the use of one versus the other.

Spatial modeling approaches

Spatial analysis is the process of learning new knowledge by taking into account the geographical context of the data. Since locations vary, the same operation may yield different results when performed in different areas. Before analysis takes place, however, it is paramount for designers to adopt an appropriate spatial modeling approach, which can have a wide array of implications on the system. To be successful, the spatial model should be able to recognize relevant features of the objects along with the processes that affect their analysis. Spatial models can take on many different flavors, as described in the points that follow:

- DeMers (2008) categorizes spatial models according to methodology. Stochastic models are based on statistical methods. Since spatio-temporal data commonly suffer from uncertainty, statistics help fill in the gaps for missing values, or generating new information from aggregated facts. Classifiers that attempt to label an event into different classes, such as SVM, nearest neighbor, or decision trees, are common examples. The converse would be deterministic methods. Instead of relying on heuristics, deterministic methods work on observed features and interactions to make a decision of interest.

- Goodchild (2003) identifies spatial models either as static or dynamic. In the former, the system accepts inputs and transforms them with predefined functions. The latter works iteratively from an initial set of conditions, outputting results at time intervals.

- Network models are hardly new, but they have increasingly come to light with the advent of social media. They often deal with the flow of information, propagation, influence, and social interaction between entities. Common uses relate to disease control, pollution detection, spread of riots and more recently, the influence of fake news.

- Spatio-temporal models not only operate on an object in its physical location and time, but also attempt to understand its behavior as it evolves (Bolstad 2005). Thus it would not suffice to detect that a riot is taking place, but it must be know its direction of propagation and side effects. Since different behaviors are elicited as locations change and time elapses, spatio-temporal models are often burdened with heavy processing tasks. While in a traditional static system an entity would have a location, in a spatio-temporal system there might be many, one location per timestamp. This modeling approach reflects current technology where information is generated and transmitted at a rapid pace by modern devices.

The decision on which model to select requires thorough analysis. It must be noted, however, that the above models are neither exhaustive, nor are they exclusionary. In fact, they may work best when combined in turn with other systems. Many modern systems, for instance, rely on spatio-temporal models, while applying features that combine quantitative analysis and iterative techniques. Other systems incorporate human feedback by

allowing the analyst to correct information or fine tune the course of direction.

References cited in Appendix C

Bolstad, Paul. 2005. *GIS Fundamentals: A First Text on Geographic Information Systems,* 2nd edition. Hamburg, Germany: Eider Press.

DeMers, Michael N. 2008. *Fundamentals of Geographical Information Systems,* 4th edition. Indianapolis, IN: Wiley.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data Via the EM Algorithm." *Journal of the Royal Statistical Society* 39(1): 1–38.

ESRI. 2001. "ArcGIS Spatial Analyst: Advanced GIS Spatial Analysis Using Raster and Vector Data." White paper. Redlands, CA: ESRI. Available: https://www.esri.com/library/whitepapers/pdfs/arcgis_spatial_analyst.pdf.

Goodchild, Michael F. 2003. "Geographic Information Science and Systems for Environmental Management." *Annual Review of Environment and Resources.* Vol. 28: 493-519. doi: https://doi.org/10.1146/annurev.energy.28.050302.105521.

Kim, Jae-On, and James Curry. 1977. "The Treatment of Missing Data in Multivariate Analysis." *Sociological Methods & Research* 6(2): 215–240. doi: https://doi.org/10.1177/004912417700600206.

Shirabe, Takeshi. 2005. "Modeling Topological Properties of a Raster Region for Spatial Optimization." In *Developments in Spatial Data Handling*, conference proceedings from 11th International Symposium on Spatial Data Handling, Peter F. Fisher, editor, 407–420. Berlin: Springer.

Appendix D: Event Harmonization Prototype

This section provides a brief overview of the event harmonization proto-type, including screenshots and description of the process. The overall event harmonization process is depicted in row 2 of Figure D1 below and is also described in the main body of the report in section 4.2.

NOTE: Portions of this appendix are not included in this unclassified publication; content removed has been noted. Request the limited distri-bution version (Volume 2) of this publication for FOUO content.

Figure D1. Process to transform event data into mission-relevant information.
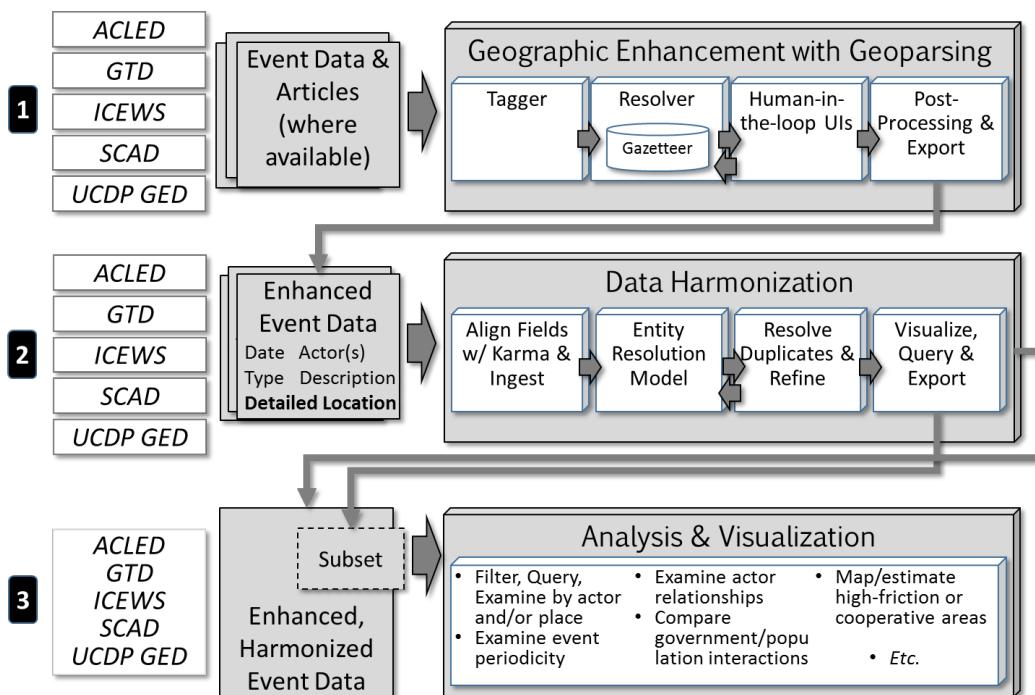
The event harmonization process assumes the data has already been col-lected, and if necessary, enhanced. Row 2 of Figure D1 depicts the follow-ing processes:

- Align fields with Karma & Ingest

  - Prepare harmonization tool for dataset ingest by loading or configuring Karma data models. Ingest datasets into data harmonization tool.

- Entity Resolution Model

    - Prepare scoring model (or reuse existing preconfigured scoring models) to detect duplicate event entries.

    - Select subset for entity resolution (if needed), such as a country (e.g., Bangladesh).

    - Execute entity resolution.

- Resolve Duplicates & Refine

    - Review entity resolution results (i.e., entries deemed duplicates); determine whether scoring model is appropriate or identifying duplicate entries incorrectly (e.g., ICEWS events that are at multiple locations in a city being deemed duplicates). Edit and refine scoring model until duplicate entries are identified, while minimizing removal of legitimately separate event entries.

- Visualize, Query & Export

    - Visualize results in harmonization tool, execute query(ies), and export results as a comma-separated-value (.csv) table, or directly ingest into analytic tool.

    - And/or, select bounding box or enter search for placename (e.g., Dhaka) and export.

Figure D2 provides a screenshot of Karma, a component used for dataset alignment within the event harmonization prototype. Karma is made available by the Information Sciences Institute of the University of Southern California.[15] Figure D2 depicts the mapping of ontology data values (i.e., Act1 – occurs_on, Act1 – has agent, etc.) to associated fields (i.e., Date_ZULU, Source) in one of the datasets (i.e., ICEWS). Karma mappings are then stored and used to align datasets to a common schema or ontology. Stable mappings may be stored as files for reuse. More details

---

[15] http://usc-isi-i2.github.io/karma/

about Karma may be found in other publications (e.g., Gupta et al. 2012; Tuchinda 2011).

Figure D2. Screenshot of Karma for data field alignment of an event dataset.
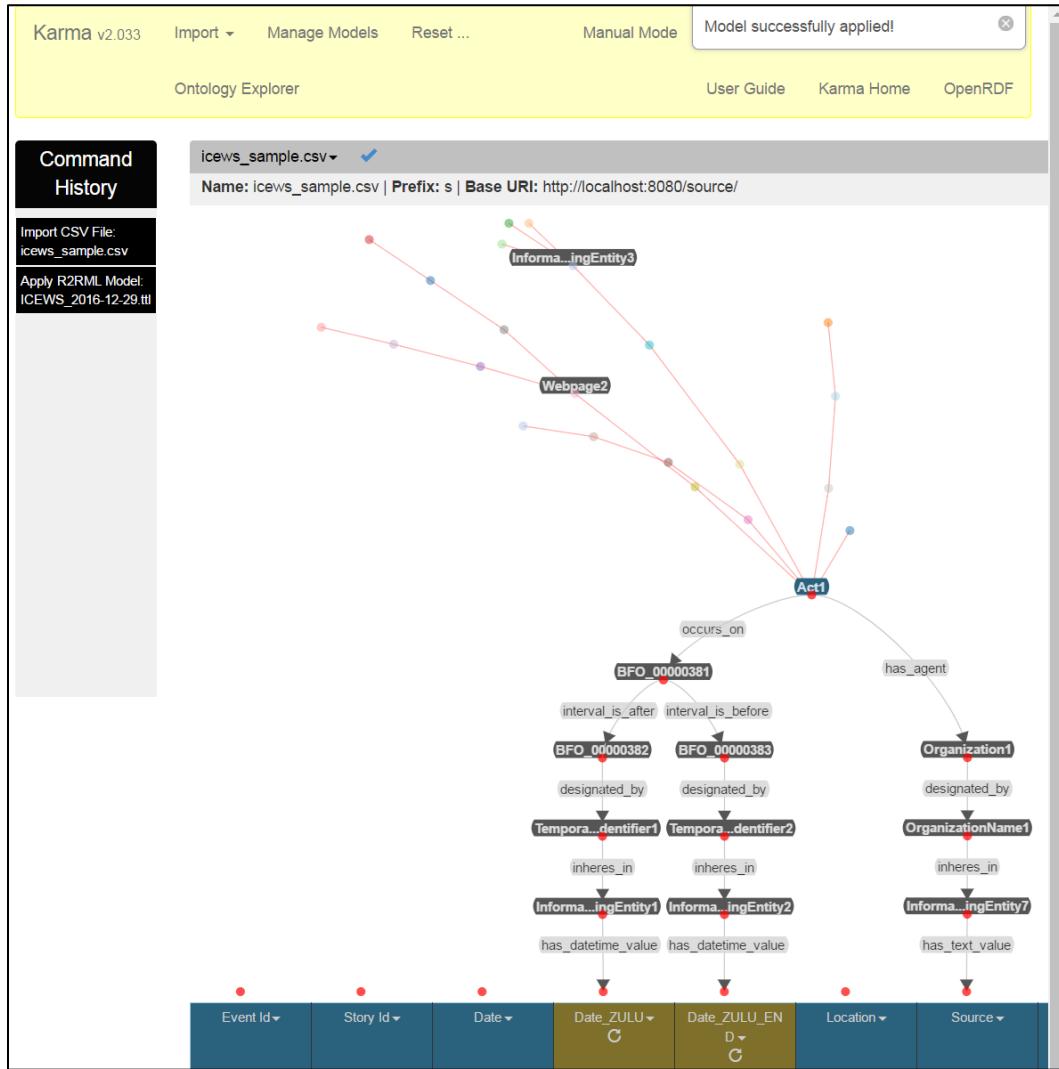
After alignment, the user uses the event harmonization prototype to complete dataset uploading or updating. Figure D3 provides a screenshot of dataset management options.

After dataset upload, the user has the option to ingest the dataset(s) into the event harmonization triple-store database, which will then allow duplicate events to be resolved and all ingested data to be returned from user searches and through the map user interface. Ingest filters may be written by using the query language known as SPARQL. Such filters may be used to reduce processing requirements to particular countries, event types, or other parameters.

Figure D3. Screenshot of interface for dataset management.

Following dataset alignment, upload, filtering and ingest, the user may configure and execute entity resolution. Entity resolution is the process that uses a scoring model to determine the likelihood that more than one entity (in this case the entity type is an event) record are referring to the same real-world thing. Configuring the scoring model is currently done through extensible markup language (XML) and requires a sophisticated user who defines the parameters used for determining matching entities. The parameters include identifying attributes (e.g., entity type, location, date or time, event type, organizations involved, descriptive text), the properties that determine whether attributes are similar, and weighting attributes to determine a similarity score. The user may then run the entity resolution scoring model, which is returned to the user as a set of results from the database, with entries associated to a new "SuperEntity" that represents a merging of individual entities in the database. The details about the original entities are retained, and the scoring model may be configured to determine which attributes are those primarily used by the SuperEntity for future analytics. The user may view the attributes of merged entities side-by-side and then qualitatively determine whether the scoring model is achieving the results expected. The user may then index the results and export the data to conduct additional analysis of the scoring model (potentially against a gold-standard dataset), or the user may begin working with the entity data.

(FOUO content removed here.)

Figure D4. *FOUO content removed, including figure.*

Scoring model results are indexed into Elasticsearch®.[16] The user may then interact with the data by using a keyword search or the geospatial interface, or the user may query the database directly by using SPARQL queries. The keyword and geospatial interface can return raw data records or filter to those that are merged entities only, and events are returned as georeferenced points on the map. The keyword and geospatial interface work as a joined query, where the bounding polygon and keyword(s) (if both are used) constrain the results together (i.e., an AND rather than OR query). The interface offers access to entity attributes in individual "baseball card" and collective tabular views. Once the user has prepared the query of interest, the user may also export data to a standard comma-separated-value (.csv) file for ingest into other tools.

The event harmonization prototype has been developed by CUBRC, Inc.,[17] in collaboration with the U.S. Army Corps of Engineers – Engineer Research and Development Center, the Army Research Laboratory, and it also derives from research and development activities initiated under the Office of the Director of National Intelligence's Intelligence Advanced Research Projects Activity (IARPA) – Knowledge Discovery and Dissemination (KDD) Program.

---

[16] Elasticsearch is a trademark of Elasticsearch BV, registered in the United States and other countries.

[17] CUBRC is a private nonprofit research and development company, headquartered in Buffalo, New York.

REPORT DOCUMENTATION PAGE

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE		3. DATES COVERED (From - To)
November 2017	Final		

4. TITLE AND SUBTITLE
Social and Political Event Data to Support Army Requirements: Volume 1

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
T41

6. AUTHOR(S)
Timothy K. Perkins, Colin D. Wood, Raimundo F. Dos Santos Jr., William D. Meyer, Noah W. Garfinkle, Xue Wang, Susan I. Enscore, Lucas A. Selig, and George W. Calfas

5d. PROJECT NUMBER
455009

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Engineer Research and Development Center (ERDC)
Construction Engineering Research Laboratory (CERL)
PO Box 9005
Champaign, IL 61826-9005

8. PERFORMING ORGANIZATION REPORT NUMBER
ERDC/CERL TR-17-40

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Assistant Secretary of the Army for
Acquisition, Logistics, and Technology
103 Army Pentagon
Washington, DC 20314-1000

10. SPONSOR/MONITOR'S ACRONYM(S)
ASA(ALT)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release. Distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Military success requires applying judgement and decision making in a high-tempo atmosphere, based on available information. Geographic da-ta at the city level is not enough spatial fidelity for tactical-level analyses. Violet Events Socio-Cultural Analysis (VESCA) work enables an analyst to evaluate and integrate multiple data sources, work with enhanced event data spatial resolution, and analyze and/or visualize the data to produce mission-relevant information. Hand-coded datasets can be more precise, but they require added time and labor to produce, have a significant lag between last observation and present day, are produced with varying schemas, and often duplicate events across datasets. This report includes background regarding event data sources; study of pro-tests, demonstrations, and rallies; and relevant analytical methods. It describes doctrine regarding civil considerations, sociocultural analysis, and contingency basing to present how event data can be transformed from its original form and interpreted to support doctrinal analysis. The report also describes enhancing event data through geoparsing and through harmonization processes and tools to align datasets to a com-mon schema and identify duplicate entries. Finally, the report presents how data may be analyzed and processed for mission-relevant results. The VESCA team's work yielded an event data harmonization prototype and recommendations for refinement.

15. SUBJECT TERMS
Geospatial data, Geographic information systems, Cities and towns, Situational awareness, Military planning, Military bases, Violent Events Socio-Cultural Analysis (VESCA)

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			**19b. TELEPHONE NUMBER (include area code)**
Unclassified	Unclassified	Unclassified	UU	94	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. 239.18